DOCUMENT RESUME

ED 333 161 CE 058 156

AUTHOR	Bhola, H. S.
TITLE	Evaluating "Literacy for Development" Projects, Programs and Campaigns. Evaluation Planning, Design and Implementation, and Utilization of Evaluation Results. UIE Handbooks and Reference Books 3.
INSTITUTION	German Foundation for International Development (DSE), Bonn (Germany).; United Nations Educational, Scientific, and Cultural Organization, Hamburg (Germany). Inst. for Education.
REPORT NO	ISBN-92-820-1059-7
PUB DATE	90
NOTE	312p.
AVAILABLE FROM	Unesco Institute for Education, Feldbrunnenstrasse 58, W-2000 Hamburg 13, Germany.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Administration; Adult Basic Education; Adult Literacy; Developing Nations; Economic Development; *Evaluation Methods; Evaluation Utilization; Evaluators; Foreign Countries; *Literacy Education; *Management Information Systems; *Naturalistic Observation; Planning; Program Evaluation; Proposal Writing; Qualitative Research; Statistical Analysis; Technical Writing; Training
IDENTIFIERS	*Naturalistic Evaluation; *Rationalistic Evaluation

ABSTRACT

This book presents a comprehensive treatment of the subject of evaluation as applied to literacy programs, covering evaluation theory, planning, and practice. Part I discusses questions of definition, context, objectives, and functions of evaluation and presents descriptions and analyses of evaluation paradigms and models. In Part II, the interrelated processes of evaluation planning and management are discussed, and evaluation planning and management approaches are explained and demonstrated. Parts III, IV, and V focus on the three components of the evaluation management approach discussed in Part II: management information systems (MIS), naturalistic evaluation (NE), and rationalistic evaluation (RE). Chapters in these parts cover: (1) theory, questions, and design of an MIS, NE, or RE; (2) writing a proposal for an MIS or for an evaluation study in the naturalistic or rationalistic mode; (3) tools and techniques of the three approaches; and (4) writing periodical and special reports. Part VI discusses the politics of evaluation, the need to establish evaluation standards for meta-evaluations, and the related question of evaluators' training. A glossary is appended. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333161

CE058158

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sutton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CE

EVALUATING "LITERACY FOR DEVELOPMENT" PROJECTS, PROGRAMS AND CAMPAIGNS

UIE Handbooks and Reference Books

1. **Handbook on Learning Strategies for Post-Literacy and Continuing Education**
   by Adama Ouane
   (1989) ISBN 92 820 1053 8

2. **Handbook on Training for Post-Literacy and Basic Education**
   by Adama Ouane, Mercy Abreu de Armengol and D.V. Sharma
   (1990) ISBN 92 820 1054 6

3. **Evaluating "Literacy for Development" Projects, Programs and Campaigns**
   by H.S. Bhola
   Joint publication with the German Foundation for International Development (DSE) - DSE Ref. 1624C/a
   (1990) ISBN 92 820 1059 7

4. **Handbook on Evaluation for Post-Literacy and Basic Education**
   In preparation

UIE Handbooks and Reference Books 3

EVALUATING "LITERACY FOR DEVELOPMENT" PROJECTS, PROGRAMS AND CAMPAIGNS

Evaluation planning,
design and implementation,
and utilization of evaluation results

H.S. Bhola

Unesco Institute for Education
German Foundation for International Development (DSE)

The Unesco Institute for Education, Hamburg, is a legally independent entity. While the programmes of the Institute are established along the lines laid down by the General Conference of Unesco, the publications of the Institute are issued under its sole responsibility; Unesco is not responsible for their contents.

The points of view, selection of facts, and opinions expressed are those of the authors and do not necessarily coincide with official positions of the Unesco Institute for Education, Hamburg.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Unesco Secretariat concerning the legal status of any country or territory, or its authorities, or concerning the delimitations of the frontiers of any country or territory.

ISBN 92 820 1059 7

© Unesco Institute for Education and German Foundation for International Development 1990

Unesco Institute for Education
Feldbrunnenstrasse 58
W-2000 Hamburg 13
Germany

German Foundation for International Development (DSE)
Hans-Böckler-Strasse 5
W-2000 Bonn 3
Germany

DSE Ref. 1624C/a

Printed by

Robert Seemann · Bramfelder Strasse 55
W-2000 Hamburg 60 · Tel. 61 89 46

TABLE OF CONTENTS

Page

Note to the Reader ix

Foreword xi

Introduction 1

PART I

Evaluation: Context, Functions and Models

1. Evaluation -- Definitions, Context, Objectives and Functions 9

2. Paradigms and Models of Evaluation 25

Section A: Two Basic Paradigms of Evaluation 27

Section B: Models of Evaluation 34

PART II

Evaluation Planning and Management

3. Evaluation Planning 53

4. Evaluation Implementation and Management 65

PART III

Management Information System (MIS)

5. MIS -- Theory, Questions and Design 77

6. Writing a Proposal for Developing a Management Information System (MIS) 92

7. The Process at a Glance: Tools and Techniques of Implementing an MIS 106

Section A: Concept Analysis 107

Section B: Writing Indicators 123

Section C: Making Tests of Achievement 127

Section D: Testing Attitudes, Observing Actions and Results 138

Section E: Data Analysis 139

8. Writing Periodical and Special Reports Based on MIS Data 143

PART IV

Evaluation in the Naturalistic Mode

9. Naturalistic Evaluation -- Theory, Questions and Design 155

10. Writing a Proposal for an Evaluation Study in the Naturalistic Mode 168

11. The Process at a Glance: Tools and Techniques of Naturalistic Evaluation 175

12. Writing Reports on Naturalistic Evaluations, and Writing Periodical Reports Naturalistically 188

PART V

Evaluation in the Rationalistic Mode

13. Rationalistic Evaluation -- Theory, Questions and Design 195

14. Writing a Proposal for an Evaluation Study in the Rationalistic Mode 210

15. The Process at a Glance: Tools and Techniques of Rationalistic Evaluation 219

Section A: Tools and Instruments 221

Section B: Data Collection 230

Secion C: Processing and Display of Data 233

Section D: Statistical Analysis of Data 254

16. Writing Reports on Rationalistic Evaluations and Promoting Utilization of Results 257

PART VI

Some Important Related Concerns

17. Politics of Evaluation, Ethics and Standards 273

18. Conducting Evaluation Training in the Third World 280

19. Conclusions 294

Glossary of Terms 297

NOTE TO THE READER

This book presents a comprehensive treatment of the subject of evaluation covering evaluation theory, planning and practice. Those with some initial familiarity with the subject of evaluation may read the material in the order in which it has been presented in the book. Others may use "random access" to read various parts of the book as appropriate.

Those responsible for planning and management of evaluation within literacy and development programs may, during their first reading, skip Chapter 2, "Paradigms and Models of Evaluation", in Part I of the book. They should read Part II of the book in full; and then go on to Part III, or Part IV, or Part V as appropriate.

Those of the practical bent, caught in the immediate need to design and conduct evaluations, may read Chapter 1, "Evaluation -- Definitions, Context, Objectives and Functions", and Chapter 3, "Evaluation Planning". They may then go on to appropriate parts or chapters in various parts as needed.

Readers are invited to send to the author their ideas and suggestions that could be used in a subsequent revision of the book to make it more useful to readers. Such suggestions will be gratefully received. Please write to:

Professor H.S. Bhola
School of Education
Indiana University
Bloomington, Indiana 47405
USA

FOREWORD

Since 1980, the Unesco Institute for Education (UIE) has been researching and promoting literacy projects, programs and campaigns with a developmental aim, in the context of its commitment to lifelong education. A major part of that effort has been to offer research-based training opportunities to educational development officers, policy-makers and practitioners in the field of post-literacy and continuing education. This training has covered the learning strategies, the training of middle and grassroots level project staff, and the evaluation of post-literacy and other nonformal basic education programs.

Through this handbook written by H.S. Bhola, UIE is able to share with a much wider audience relevant processes and strategies for evaluating literacy for development programs. The first draft of Prof. Bhola's book was commissioned by the German Foundation for International Development (DSE), and we welcome the initiative taken by this institution in inviting UIE to finalize and publish this revised version.

Instead of sets of procedures that can be applied indiscriminately to all situations, we find here the questions and guidelines which need to be considered in order to design realistic, appropriate strategies for formative and summative evaluation. Not only the various theoretical models are covered but also, above all, the practical aspects of evaluation, with examples of case studies, questionnaires, instruments and processes drawn both from the author's own extensive and widely acknowledged experience in the field, and from other actual projects.

The book addresses the evaluation of learning outcomes, of project management structures and of in-service training. At the end, the broader context of educational development is considered, as evaluation does not take place in a socio-political vacuum. Achievements are measured against objectives, which are themselves a reflection and a part of political decision-making.

Referring to the on-going methodological debate, Professor Bhola rightly refuses to choose exclusively qualitative or quantitative approaches. The key issue is the need to shift from a positivistic paradigm to what he calls a "naturalistic" approach, which accords with UIE's longstanding research methodology. This is not to say that measurement is obsolete, but that the choice of appropriate approaches and methodologies (quantitative and qualitative) is bound to be related to the type of information needed, to the context and to the evaluation issues at stake. A crucial issue is then the integration of the evaluation process within the overall education program.

This book is a practical one that intends to help planners, trainers and field workers to introduce internal evaluation processes into their programs, to plan the production of the kind of information needed to take better decisions for the improvement of programs, to monitor the implementation of objectives, and to understand and improve the existing informal evaluation processes. Indeed the development of evaluation is fundamentally the development of a "learning culture" within an organization or a community.

We are indebted to Professor Bhola for his painstaking revision of his original draft, while without the warm cooperation of the DSE this publication would not have been possible. I express thanks in particular to Dr Joseph Müller of the DSE for his support and advice. Within our Institute, I am particularly grateful to Dr Adama Ouane, the coordinator of UIE's studies on evaluation pertaining to literacy, post-literacy and nonformal education, and to Mr Peter Sutton, Ms Wilma Gramkow and Ms Dietlind Oschlies, who have supervised the publication and edited the text.

Paul Bélanger
Director
Unesco Institute for Education

INTRODUCTION

This is a book on evaluation, surely an important social concern today. Evaluation has become important for reasons both professional and political. Professional planners and managers want evaluation to ensure better implementation and, thereby, greater effectiveness of their programs. Politicians demand evaluations to promote greater accountability all round. Evaluations of literacy projects, programs and campaigns are being required for the same set of reasons.

If literacy is a Human Right, then "literacy for literacy's sake" is justified. Human rights and human fulfilments should not have to be justified for any extrinsic reasons. In the real world, however, literacy has still to be justified on the criteria of *functionality*. In the 1960s and 1970s, the criteria of functionality were narrowly economic. Fortunately, during the 1980s, the concept of functionality was expanded to include the economic *and* the political, social, educational, cultural and environmental. In this book, we accept literacy as a human right but at the same time as more than a mere ideological ornament. Literacy is seen as "potential added" to individual capacities and collective possibilities. Literacy is seen as enabling individuals to make more effective transactions with all aspects of their environment -- economic, political, social, educational and cultural. Literacy is seen as "symbolic capital" that nations must join with "material capital" to bring both democratization and modernization to their peoples.

While the book is addressed directly to literacy workers, it is a book for *all* development workers. It should be of interest to educators working both in formal and nonformal settings, to agricultural extension workers, health educators, family planners, and cooperators.

In this book we have talked of evaluating "literacy for development" -- projects, programs and campaigns. These three approaches to the delivery of literacy services are indeed different approaches as far as the politics of literacy is concerned. Each of the three approaches involves a different level of political commitment, and a different style of mobilization of peoples and resources. However, in regard to the evaluation of literacy projects, programs and campaigns, there are no significant differences. Therefore, evaluation planning, management and implementation for all the three modes of delivery of literacy services remain the same, except for some differences in scope and style.

In this book we have taken a set of definite professional positions. We have come to these positions on the basis of experience of conducting evaluation workshops for literacy workers and development agents in Latin America, Asia and Africa -- particularly in a series of workshops in Tanzania, Kenya, Botswana, Malawi and Zimbabwe during 1979-1989.

First, we are committed to the ideology of *internal* evaluation. Literacy workers and development agents, we believe, should pay serious attention to internal evaluations. Such evaluations absorb relatively fewer resources and can be used immediately to improve the effectiveness of literacy projects, programs, and campaigns. Literacy workers should leave *external* evaluations to outsiders, who may do them to fulfill their own special policy and political needs. Indeed, if literacy and development workers have conducted their own internal evaluations, they will be better able to collaborate with external evaluators.

Second, we have come to take the position that "information" is the *master concept* in planning and implementation, and not necessarily "evaluation". We believe that what we need for effective implementation is information for decision-making; and that all information does not necessarily have to be generated through specially designed evaluation studies. Useful data are routinely generated by "literacy for development" campaigns, programs and projects, in the very process of their planning and implementation. When systematically collected and stored for later retrieval and use, these data would constitute a Management Information System (MIS). We have realized that such MIS data and periodical reports written by staff, do indeed constitute the information most used by decision-makers in their day-to-day decisions. It is for this reason that the design, installation and utilization of an MIS is now an important part of the book; and is presented as the cornerstone of any evaluation planning and management. We must hasten to add that these MIS's can be paper-and-pencil systems and need not wait for computer technology.

Third, we have realized the immense usefulness of what are called naturalistic inquiry approaches to evaluation. We have learned that quite often the search for the so-called scientific and objective evaluation was no more than an exercise in "scienticism" without making much "good sense". We are convinced that it is not a weakness but a merit for an evaluation to be contextual, responsive and qualitative, as we try to put "a frame on the flux" of field realities. Naturalistic evaluation, we have found out, could be scientific and systematic; and *in its own terms*, it could be objective, reliable and valid. The terms more relevant to the study of human actions were consistency, coherence and credibility. As evaluators, we had to be able to make "warranted assertions" within a "network of plausibilities" rather than within a "network of causalities". Indeed, there are questions which only evaluation in the naturalistic mode could tackle and answer. The present edition of the book brings the discussion of naturalistic evaluation right to the center of the book, it is not merely a tack-on as it was in an earlier edition. Yet, we do not suggest that evaluators should stop counting! Nor do we reject what we have called rationalistic evaluation (RE). What we do suggest is that RE should be used only when it is best able to answer the evaluation question on the evaluation agenda.

Fourth, the author has discovered the necessity of the process of "evaluation planning". There is a lot of talk in the planning literature of "development" planning and "educational" planning, but the phrase "evaluation planning" does not occur too often, if at all. We have realized that it is important for both internal and external evaluators first to take a comprehensive look at all their information needs and then to develop an evaluation agenda. Such an agenda should respond, on the one hand, to high-priority information needs, and, on the other hand, should be sensitive to existing resource constraints. Consequently, this edition of the book pays due attention to the concept and process of evaluation planning. This may also be one of the first books to give due attention to the process of "evaluation management", presenting a particular model for generating evaluative information for use in the management of literacy and development programs.

This is a book for all evaluators everywhere in the world but it is of special usefulness to planners, trainers and field workers in the Third World. This fact has determined the general approach, the content, and the level of discussion presented in the book. For some development agents in the Third World, this may be the very *first* book they may be reading on the subject of evaluation. Therefore, considerable attention has been paid to the choice of language as well as to the organization of the content in the book. We have tried to be clear and simple, without being simple-minded. Discussion of evaluation does involve technical vocabulary that had to be introduced to the reader. However, a glossary of terms has been appended to the book to enable readers to master the "language of discourse" in the area of evaluation. Content has been so organized that readers can see through the argument presented and get to the point. Often, important points have been highlighted in the form of numbered lists rather than in running paragraphs. A set of charts spread throughout the book summarizes the total argument in graphics.

In addition to being the first book ever on evaluation to be read, for some development workers and trainers this may be the *only* book that they are able to obtain where they work. Therefore, this book has been made as self-contained as possible. It covers the whole range of topics from evaluation theory to evaluation practice. On the one hand, it introduces the reader to the theory of evaluation and to the politics of evaluation within organizations and communities. On the other hand, it encourages the reader to do something practical with the concepts and techniques presented in the book.

It is our hope that the book will enable literacy and development workers to conduct small-scale evaluations of their work on their own. It will be most useful if readers are first introduced to the material in the book in a workshop setting. It is not inconceivable, however, for an intelligent literacy or development worker to follow and use the book in conducting a small-scale evaluation study without too much outside help.

The idea of the publication of this book in this form took shape at the World Conference on Education for All held during March 5-9, 1990 in Jomtien, Thailand, where delegates from almost all the world's nations met and resolved to work towards the universalization of adult literacy and primary education by the year 2000. An important theme of the conference was that the success of the decade of "Education for All" will have to be judged by the results of national policies as they appear in the lives of nations, and by the consequences of knowledge acquisition on the lives of children, youth and adults.

We sincerely hope that this humble contribution will be of some use in evaluating results and consequences of projects, programs and campaigns of literacy for development and education for all in the last ten years of this century.

H.S. Bhola
Indiana University
Bloomington, Indiana
USA

Part I

Evaluation: Context, Functions and Models

This Part of the book discusses questions of definition, context, objectives and functions of evaluation; and presents descriptions and analyses of evaluation paradigms and models. It is divided into the following chapters:

1. Evaluation -- Definitions, Context, Objectives and Functions, and

2. Paradigms and Models of Evaluation.

CHAPTER 1

EVALUATION -- DEFINITIONS, CONTEXT, OBJECTIVES AND FUNCTIONS

Evaluation is a process of judging the merit or worth of something. As human beings, we engage in the process of common-sense evaluation all the time. Professional evaluation, however, is more than common sense. It is sensibly organized; it is as precise as possible; and its results are both warranted and publicly defensible. The essential objective of doing professional evaluation is to generate information that can be used in the planning and implementation of programs to improve the quality of life. Evaluation may take many forms, for example, needs assessment, base-line survey, learner evaluation, personnel evaluation, achievement and attitude testing, curriculum evaluation, analysis of organizational capacity, product evaluation, assessment of impact, cost-benefit analysis, self-evaluation, and others. Evaluation has come to acquire functions that go beyond the informational. It often serves functions that are institutional, social, historical and political.

Common-sense evaluation

The word *value* is built right into the word "e*valu*ation". Indeed, evaluation means assigning values to judge the amount, degree, condition, worth, quality or effectiveness of something.

As human beings, we are perpetual evaluators. We evaluate things as we go shopping. We evaluate people as we choose friends, spouses and workers. We evaluate books to read and films to see. We evaluate bars and restaurants as we make plans for the evening. We evaluate our personal and official actions and their effects. We evaluate communities and environments as we make decisions about buying or renting a home or choosing schools for our children. We evaluate party manifestoes and sincerity of leaders and cast our votes on that basis.

Evaluation, as we have talked of it so far, is a *personal act, and* it often lies in the personal domain. We can be more or less self-conscious and more or less cautious about our personal evaluations, but these evaluations remain impressionistic. These are *common-sense* evaluations.

Common sense must continue to take a central position in what we might call *professional* evaluation. But in professional evaluation, we do go beyond mere common sense. With professional evaluation, we acquire a social context as we come into the institutional and the public domain. We are acting in behalf of development institutions, spending public funds and we are accountable to the people. Our evaluations should be able to make warranted assertions and have to be publicly defensible.

In recent years, evaluation has emerged as an area of specialization that teaches us how to be most *perceptive* and most *logical* at the same time. It has taught us a lot about how to develop descriptions, make judgements and write recommendations that are defensible.

What is evaluation?

With what we have read above, we can, of course, make our own definitions of evaluation. Here are some examples from the published literature on evaluation:

Egon G. Guba and Yvonna S. Lincoln have defined evaluation "as the process of describing an *evaluand* [the entity being evaluated] and judging its merit and worth." Merit means the inherent goodness of something, while worth means the comparative usefulness of something to somebody in a particular context.

Daniel L. Stufflebeam defined evaluation as "the process of delineating, obtaining, and providing useful information for judging decision alternatives." Marvin C. Alkin describes evaluation as the "process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives." Lee J. Cronbach defines evaluation simply as "the collection and use of information to make decisions about an educational program."

A recent book on evaluation[1] recommends that we accept the following definition of evaluation: "An evaluation study is one that is designed and conducted to assist some audience to judge and improve the worth of some educational object."

As we can see, there are some common themes in the above definitions that can be underlined. Evaluation must generate information. This information must be defensible. There should be a method to its collection. Thus, evaluation should be organized. As far as possible, information should have the quality of being exact and precise. Most importantly, the information must be usable in the improvement of some developmental, educational or training program. This orientation of collecting "information for decisions" is the most characteristic of evaluation theory today and its most note-worthy feature.

Some further definitions and differentiations should be discussed here. We begin with a distinction between evaluation and research.

Evaluation and research distinguished

Evaluation and research are two different professional activities, though the two often get confused. Confusion occurs because the evaluator and the researcher use similar inquiry designs, methodologies, tools and instruments, and have similar concerns for the defensibility of their findings. Quite often the same person may be acting as an evaluator on one literacy project and as a researcher in another setting. Evaluation and research, however, differ significantly in terms of their inquiry frameworks and their task objectives. The table on page 12 should clarify the distinctions.

Evaluation and supervision

Another useful distinction can be made between evaluation and supervision. *Supervision* itself is difficult to define. At its worst, supervision is equated with watchdog functions of those above watching over those below. At its best, supervision is seen as an educational process wherein the more experienced colleagues mentor those relatively new to their jobs and support those who are able to function autonomously. On the other hand, *evaluation* need not always be negatively judgemental. Thus understood, good supervision should enable the various functionaries in a program to analyze and evaluate their own performance in relation to the program needs, and to learn and grow on the job.

Indeed, the distinction between evaluation and supervision is breaking down. Supervision is incorporating evaluation strategies.

Evaluator	Researcher
Policy and planning orientation; seeks to clarify planning alternatives and to improve program performance.	Disciplinary and academic orientation; seeks to advance the frontiers of knowledge in the researcher's own discipline.
Loyalty is to a particular literacy campaign, program or project; choice of evaluation topics is determined by the information needs of decision-makers.	Loyalty is to a particular academic discipline; choice of research topics is determined by the theory and research needs of the discipline.
The methodological choices are "scientific" but non-experimental and quite often naturalistic. The norm for judging the findings is applicability to the program situation and adaptability to other similar program settings.	The methodological choices are "scientific" but often there may be emphasis on control. Experimental or quasi-experimental methods may be preferred. The norm for judging the findings is generalizability or transferability.
Time-frame for the production of results is set by the program.	Time-frame for the production of results is set by the researcher and by the internal logic of the research question.
Professional rewards consist in the utilization of findings by decision-makers and demonstrated improvement in program implementation.	Professional rewards consist in publication of findings in professional journals and favorable comments by professional colleagues.

In this monograph, we shall take the position that supervision cannot in fact be separated from evaluation; and that indeed evaluation can and should contribute to the individual growth of those being supervised. We shall further suggest that each supervisory visit to the field should become an opportunity for naturalistic evaluation of some aspect of the program in question.

Categories and kinds of evaluation

Distinguishing between evaluation and research, or between evaluation and supervision, has not solved the definitional problems in the area of evaluation. Numerous categories, types and kinds of evaluation have been proposed and promoted. That does not necessarily help but in fact complicates the lives of policy-makers and practitioners.

Some categorizations of evaluation are highly theoretical and are presented in literature as evaluation paradigms and models. We shall discuss some of those paradigms and models in the next chapter. Other categorizations are rooted in values -- internal evaluation or external evaluation; and controlled versus participatory evaluation. Some evaluations use pragmatic and not so pure categories of resource allocations of time and effort, e.g., monitoring and quick appraisals. Some distinguish process from product -- formative evaluation versus summative evaluation.

Some distinguish among the units of analysis -- learner evaluation, group evaluation, program evaluation and evaluation of impact on communities or sub-cultures. Finally, there is a whole series of evaluation labels derived from evaluation objectives, evaluation tasks, or from what is being evaluated. Overlap among these various categories is considerable.

The categorization of evaluation used in this book can be best described as theoretical. We define evaluation as a process of *information generation.* This process is seen to consist of three approaches to information gathering: (1) operational information -- typically numerical -- generated in the very process of implementation of the program; (2) experiential information -- typically collected through naturalistic strategies -- informing us about how participants in a program are experiencing the program, and about its meaning in their real lives; and (3) comparative and correlational information, generated through more rationalistic evaluation strategies.

Monitoring and quick appraisals

There has been a considerable interest recently in development literature on strategies for gathering quick evaluative feedback on the performance of programs. The point is made that typical evaluation studies may too often take too long a time for decision-makers to wait for results of studies. Program decisions will often demand quick pulse-taking of a program to get a report card on the general health of a program. Timeliness is important. As a response to this need, evaluators have developed approaches described as monitoring and quick appraisals.

To *monitor*, in dictionary meanings, is to watch, observe, check and sometimes adjust. In evaluation literature, to monitor is, indeed, to check upon an on-going program for flaws or breakdowns to enable decision-makers to regulate activities and to undertake corrective actions. Monitoring is, thus, an important aspect of evaluation. Monitoring, however, is only a part of evaluation, and not the whole of it. Monitoring looks at performance data, routinely generated by the program in the process of implementation, and cautions the decision-makers about the gaps between expectation and reality. Monitoring is thus a systematic check on the progress of a project in the framework of original goals, procedures and results - a sort of *performance audit*. Good monitoring typically requires a Management Information System (MIS) to support it by keeping records of inputs and outputs and other related indicators. We shall have a lot more to say about Management Information Systems (MIS's) later in the book.

Quick appraisals are quick evaluations, conducted under conditions of emergency to investigate the cause of a breakdown, to anticipate problems, or to get early returns on the impact of a program. Quick appraisal is the child of necessity. It is undertaken when there is no time to wait for a regular evaluation.

A quick appraisal will use monitoring data routinely generated by the program in the process of planning and implementation as well as secondary data from other sources. Quick appraisals will, additionally, collect fresh data for the purpose of answering significant questions. Such data may, for example, include self-reports by functionaries of the program.

Quick appraisals may be less exhaustive and less comprehensive than regular evaluation studies, but one needs to prepare for a quick appraisal carefully and systematically. Appraisal teams will have to be carefully built, and given a clear mandate regarding the information they should collect and the judgements they should render. The team should think about the mix of quantitative and qualitative data they should try to collect. Instruments should be carefully designed and tested. Samples should be small but, again, carefully chosen. Deadlines should be met, otherwise a quick appraisal is no longer a quick appraisal. A time-frame of four to six weeks is typical for quick appraisals.

Internal evaluation versus external evaluation

The question of internal versus external evaluation is really the question as to who will conduct an evaluation. The dimensions of *externality* can be many: international donor versus recipient nation; agent from one institution evaluating program of another institution; one unit evaluating another unit within the same institution; and an evaluation specialist evaluating the work of a program specialist within the same one unit.

In our definition, internal evaluation means that the program people do their evaluation themselves, and even when they use a specialist as a consultant they are in control of the process of evaluation -- in formulating questions, in the choice of methods, design of study, data collection and analysis, establishing criteria for success of program, and use of evaluative information for future planning. The process or results of evaluation are not kept secret from anyone except to protect the innocent. Thus, internal evaluation is that conducted within the program system by program specialists themselves. External evaluation is that conducted by evaluators sent from outside the system.

It is often asserted that external evaluation is more objective than internal evaluation, which is then rejected as being both subjective and political. On the other hand, program specialists often dread external evaluations, which they complain are "parachute evaluations" -- often hurried, superficial, uninformed and sometimes clearly political.

It should be pointed out that external evaluations are by no means inherently objective; and internal evaluations are by no means naturally lacking in credibility. Both internal and external evaluations can be highly political, and, therefore, suspect. On the other hand, both internal and external evaluations can serve important purposes in the context of special needs. There may be instances when external evaluation is necessary for making policy and planning decisions at some levels of decision-making. Most of the time, however, it is internal evaluation that makes sense, enabling practitioners to take control of their program and helping them to grow in the process.

Participative evaluation, collaborative evaluation, collective evaluation

People-centered values of our times have led to the conceptualization and implementation of what has been come to be called participative evaluation. Essentially, participative (or participatory) evaluation means that it is not evaluation done by an outside expert in splendid isolation from the people, wrapped within the pretense of objectivity, but an evaluation that is done by all the stakeholders concerned, together, in participation with each other. All those involved, and particularly, the learners and participants in program activities, together construct their own meanings, and speak in their own behalf, in their own language. Collaborative evaluation is another name for a somewhat similar approach to evaluation. Finally, collective evaluations of processes and events have been attempted by large collectivities, in seminars or in large groups assembled in halls and stadia, providing testimony on how a program functioned and how it was experienced by the people in their day-to-day lives.

Formative evaluation versus summative evaluation

The concepts of "formative" evaluation and "summative" evaluation, introduced by Michael Scriven, have come to be two of the most commonly used concepts in the discussion of evaluation. Both these concepts are simple to understand. Formative evaluation is evaluation of a curricular product or a program in the very process of its formation. The emphasis is on process. The information generated can be used in improving the curriculum-in-the-making or the program during its implementation. Summative evaluation is to sum things up. It comes at the end of a literacy program or at the end of a curriculum development phase within the program.

Objectives-related and task-related forms of evaluation

The objectives of evaluation, as we have indicated above, are always informational. Evaluators strive to *collect usable information.* But one may need usable information on the *context* of the program, on the quality of *inputs* made into the program, on the *processes* of instruction and organization, or on the *outputs* and *outcomes.* Again, the intention may be to modify and improve, to compare or contrast, or to make decisions about the continuation or termination of programs of development or training.

In the following, we first list, and then describe very briefly, the various objectives- or task-related forms that evaluation might take:

1. Needs assessment
2. Base-line survey
3. Learner evaluation
4. Achievement and attitude testing
5. Personnel evaluation
6. Curriculum evaluation
7. Institutional or organizational evaluation
8. Product evaluation
9. Impact evaluation
10. Cost-effectiveness evaluation, and
11. Self-evaluation

1. *Needs assessment*

Evaluators may have to conduct needs assessments at various levels of the system. They may conduct a general needs assessment at the national level to reflect those needs in the design of the "literacy for development" campaign, program or project. They may also do a needs assessment at the community level to see what demands a particular literacy campaign, program or project will make on literacy workers and development agents. Finally, needs assessment may be conducted within groups of learners and trainees to select teaching content and to design appropriate teaching strategies. A good needs assessment will typically cover all the constituencies involved within the relevant system -- adult learners, facilitators, trainers, field supervisors, administrators in education and extension departments, community leaders, and members of communities themselves. The final program design should be done on the basis of the various needs profiles generated by these different constituencies and groups brought together through a process of honest needs negotiation.

2. *Base-line survey*

Base-line surveys of communities are undertaken to establish the economic, social and cultural base-line against which later changes can be judged. Community development and literacy workers generally would conduct extensive base-line surveys in communities they seek to serve. Wherever possible, evaluators should use already available base-line data to design their programs for literacy workers and development agents. It is possible, however, that the base-line survey already conducted had not anticipated the special information needs of literacy workers or their trainers. In that case a new base-line survey would be defensible, even necessary.

To take one specific example: The special information needs of trainers-evaluators may deal with (1) role considerations, and (2) knowledge considerations. A trainer-evaluator preparing family health education workers, for example, would need to know the current child-rearing and health practices within communities; level of knowledge of nutrition; lack or otherwise of home gardening; and level of consumption of animal proteins. At the same time, the trainer-evaluator would be interested in how this knowledge is currently acquired by mothers; whether traditional educational roles exist that disseminate this information; what other more modern secular roles have already been introduced within those communities by the government; and what expectations one should have about the introduction of a new role of the family health education worker.

As can be surmised, "literacy for development" workers will have to design base-line surveys to fit their special information needs about existing literacy levels, information seeking patterns, development and education roles in the community, and levels of development knowledge in the community.

3. *Learner evaluation*

Literacy workers had for long resisted learner evaluation. Their position was that adult men and women who came to attend literacy classes should not be insulted by being subjected to "examinations" and then humiliated by being placed in pass and fail categories.

Pragmatic reasons were cited as well. It was argued that the motivations of those who came to literacy classes were already so low that they would use any excuse for dropping out of the program, and the terror of the test would surely provide them with the excuse to do so.

Donor agencies that have provided grants and loans for literacy work to the Third World have persistently insisted that learner evaluation be conducted to assess *results* of efforts. Most funded literacy programs are now being obliged to conduct learner evaluations. Most often learner evaluation ends up being a combination of achievement tests and attitudinal testing. Attempts are being made to make these learner evaluations as little threatening to adults as humanly possible.

4. *Achievement and attitude testing*

There is considerable overlap between the conception of "learner evaluation" and "achievement and attitude testing." Much can be learned by testing learners' achievement of knowledge in agriculture, health, and cooperation; and by testing their attitudes towards family planning and national integration. A considerable part of evaluation within a literacy system will consist of testing. It will be testing of learners as they enter the literacy project and their testing as they leave. This testing will cover knowledge they have learned; their diagnostic and performance skills; their motivations, attitudes and values; and their communication and production skills. Some of this testing may have to be done not on the learners themselves but on other individuals in the communities, to be able to record the dissemination of new knowledge and the filtering of new attitudes within communities.

5. *Personnel evaluation*

Personnel evaluation involves the assessment of the competence and commitment of functionaries in a program: How good are the planners, the administrators, the program specialists, the trainers, the teachers, and the field workers?

In most developing countries of the world -- as also in the developed countries -- personnel policies are such that people once employed cannot be let go too easily. Once aboard, they can neither be dismissed nor transferred. In such cases, personnel evaluation should be done for the purposes of "staff development".

6. *Curriculum evaluation*

We seldom think of curriculum in relation to "literacy for development" programs. But literacy for development programs do have a curriculum in the general sense of a "course of study" and a "course of action". In the context of the evaluation of literacy instruction, curriculum will be a frequent evaluation theme of evaluators. They may need to evaluate particular items of instructional materials -- a primer, a handbook, a set of charts, a simulation game. They may want to evaluate a particular teaching or training method, for example, team facilitation versus single tutor. Different systems of program delivery may be tested: correspondence courses versus night schools; teaching individual learners or teaching families, etc. Finally, the overall effectiveness of a "literacy for development" curriculum may be the concern of evaluators.

7. *Institutional or organizational evaluation*

The quality of institutions or organizations determines the quality of services these organizations will be able to produce and deliver. Unfortunately, very little attention seems generally to have been paid by literacy workers and development specialists to organizational traits. Institutions or organizations can be studied along two general dimensions: (1) organizational climate, and (2) organizational capacity. Organizational capacity is determined through an accounting of an organization's resources in relation to its mission. Organizational climate is a conceptualization of an organization's social life -- members' identification with the organization and their satisfaction or dissatisfaction with the organization's decision-making style and patterns.

8. *Product evaluation*

Programs of development or literacy teaching produce various products: primers, follow-up books, teacher manuals, trained personnel, and local institutions, such as community centers, cooperatives and banks. Some of these products will be evaluated as part of curriculum evaluation; and some under personnel evaluation. However, product evaluation is a concern that deserves special evaluation and requires different strategies depending upon the product in question which may be a book, a film or an object crafted for income generation.

9. *Impact evaluation*
The study of the impact of literacy or post-literacy initiatives on beneficiaries must go beyond the testing of curricula, materials and learners within organized instructional settings. Evaluators must go into the communities in which their learners live and work. Their questions must, however, be sharply focused: Did the literacy teacher as a change agent fit into the social setting? Was the new role performer able to teach, demonstrate and resocialize? Did learners learn? Did community development occur as a result? These questions can be answered if base-line data were collected earlier.

The study of the impact on communities must provide proper time for the new ideas to go through the period of adaptation and use by the communities. They should have time to relate, learn and adopt. Such "sink-in periods" may have to be many months (if not many years) long. Also, in the study of the impact of new roles and new teaching within communities, evaluators should look for both the anticipated and the unanticipated consequences of the introduction of new change agents, new learning and new attitudes. Have role conflicts emerged in relation to traditional roles? Is a new group of power holders emerging within communities because of new roles to be performed? Has the change agent brought in bureaucratization resulting in the destruction of local initiatives? Are learners putting their learning to work in their daily lives?

10. *Cost-effectiveness evaluation*
Two terms are in use in the literature of evaluation in the development sector: cost-benefit analysis and cost-effectiveness analysis. Both analyses involve comparisons of costs and outcomes, but the nature of comparisons differs.

Cost-benefit analysis is possible when outcomes can be given clear economic values in dollars and cents. This sort of economic analysis is seldom possible in education and extension, where non-material effects are the most significant but cannot be assigned numerical values. Cost-effectiveness analysis is used where outcomes cannot be expressed in monetary terms because of the absence of market prices for outcomes. Therefore, the levels of outcomes themselves are compared in proportion to the costs incurred in each different case.

11. *Self-evaluation*
Learning to evaluate is professional growth. To engage in self-evaluation is growth in both the professional and the personal sense, and at much deeper levels. At its simplest, self-evaluation is introspection. This introspection can proceed along both the dimension of value clarification, and that of analysis of discrepancies among what was expected, what was possible, and what was actually achieved. This analysis can be more than impressionistic and can be based on notes and records.

Functions of evaluation

Functions of evaluation go beyond its typically stated objectives. Its objectives generally are informational, but its functions are, at the same time, informational, professional, social-psychological, political and historical.

Informational: The informational functions of evaluation are quite obvious. These are to provide feedback and to create usable information -- information that can be utilized to improve on-going programs.

Professional: The professional functions of evaluation are to increase understanding about the means and ends of a program; to demonstrate the effectiveness or failure of plans and strategies in use; and to suggest corrective actions. It is important to note that evaluations are conducted not merely to find faults with a program, but also to demonstrate its strengths and goodness.

Organizational: Evaluation fulfills important organizational functions. At its best, it helps organizations to undertake organizational renewal by forcing an examination of goals and purposes, reducing bureaucratic complacency, and clarifying standard operational procedures buried under day-to-day routines.

Political: The political functions of evaluation include agenda setting and generating debate on important issues. It promotes accountability, and can promote citizen participation. On the one hand, evaluation can legitimize an on-going program and on the other hand, it can look for scapegoats to fix blame, and can kill the programs which the political actors may have decided to terminate in the first place. Evaluation can perform some radical political functions as well by promoting the interests of clients and constituencies that otherwise would never have had a voice.

Social-psychological: Evaluation's social-psychological functions can be those of pacification and mystification -- to give clients and citizens a feeling of security, by reducing complex social problems to a choice between relatively simple alternatives. In its more positive aspects it can promote conflict resolution and arbitration.

Historical: Finally, evaluation has important historical functions -- to record and to document actions, events and results that otherwise might be lost to collective memory.

In the next chapter, we shall discuss the two major paradigms of evaluation and research as well as the various models of evaluation that have been proposed during the last twenty-five to thirty years.

Things to do or think about

1. What was your definition of evaluation before you read this chapter? How has it changed now?

2. Did you have the opportunity of being somehow associated with an external evaluation or an internal evaluation in your professional life? What were your experiences? Were results from these evaluations utilized by decision-makers or anybody else?

3. Are you clear about the distinction between evaluation and research? Try to explain the difference between the two to a colleague to his or her satisfaction.

4. In the chapter you have just read, many different forms of evaluation have been discussed. Has your department conducted or participated in the conduct of one or more of these forms of evaluation? Which one(s)? With what consequences?

5. Do you have a "story" to tell about the political functions of evaluation?

Note

1. Madaus, George F. *et al.*, eds. *Evaluation models: viewpoints on educational and human services evaluation.* Boston, MA: Kluwer-Nijhoff, 1983.

CHAPTER 2

PARADIGMS AND MODELS OF EVALUATION

A model is the essence of a model-maker's professional experience. This essence is itself formed within the framework of the model-maker's particular view of "how the world works". Since scholars and practitioners of evaluation differ from each other in their world-views and in their professional experiences, many different models of evaluation have been proposed, among them the CIPP model, the discrepancy model, the transactional model, the goal-free model, the investigative model, evaluation as illumination, the connoisseurship model, the advocacy model, and the participative model. This is by no means a complete list of all the models available in the literature.

All these various models (approaches, or strategies) can be divided into two groups in terms of their governing paradigms, that is, their philosophic positions and creative ideologies. These two paradigms have been called the rationalistic paradigm, and the naturalistic paradigm. Within the two major groups of evaluation models, however, there are considerable overlaps in terms of paradigms.

In Chapter 1 (Evaluation -- Definitions, Context, Objectives and Functions), we distinguished between common-sense evaluation and professional evaluation. Professional evaluation itself can be conducted at various levels of understanding and sophistication. The use of the experimental design and highly sophisticated statistical techniques, does not, however, guarantee good evaluations. Ironically, the more systematic and formulistic the methods and design, the greater is the possibility of their being used "thoughtlessly" in evaluations!

To do evaluations that add to the understandings of both the evaluators doing the evaluation, and decision-makers using evaluation results, one has to know the *theory* of evaluation. Then, the theoretical development itself has to be put in perspective, through an understanding of the *history* of theoretical development in the field. This chapter seeks to provide the reader with a history-theory of the field of evaluation. The various models of evaluation proposed during the last thirty years or more are discussed both in their theoretical (i.e., paradigmatic) and historical contexts.

The "Note to the Reader" in front of the book advises the beginning student of evaluation to save this conceptually loaded chapter for a later time. Those who want to try reading it now should keep in mind the general idea that the various models of evaluation will be found to be leaning on the side of either the positivist or what we have called the naturalistic paradigm. Of course, there will often be conceptual overlaps and repetitions of history.

The reader should look at these models as milestones on the road of theoretical development in the field of evaluation.

Some will demonstrate the beginnings of ideas that have now become conventional wisdom. Others will illuminate the underpinnings of methods and techniques that are now in frequent use. Yet others may sow the seeds of doubt in our minds about things of which we were all so certain.

In trying to *use* these models, one need not seek to use each in full, and in pure form, by itself. Various mixes of models may be tried. Indeed, most uses of models may consist of no more than borrowing "the language" employed in presenting the models. To restate, paradigms and models are the subject of discussion in this chapter. However, to make sense of the discussion of evaluation models and paradigms of inquiry, we must first have an understanding of the terms "paradigm" and "model".

What is a paradigm?

Let us begin with the word "paradigm". Its dictionary meanings are pretentious. A paradigm is defined as an ordered list, a table of classes, a pattern, or a formula for the general form into which specifics of a certain order may be placed. In formal terms, a paradigm has been defined as an axiomatic system with a particular set of assumptions about phenomena into which it is supposed to inquire.

Kuhn[1] in his study of scientific revolutions defined a paradigm somewhat colorfully as the *creative ideology* of scientists from which they worked, and which provided them with a particular logical and methodological stance for producing scientific or social-scientific knowledge.

Thus, evaluation paradigms are the creative ideologies of evaluators. These paradigms determine the thinking and methodological behaviors of evaluators: what they think about the nature of reality; and how they think "warranted assertions", that is, trustworthy statements, can be made about the social reality that surrounds us. There are two basic paradigms of evaluation (and research) that we shall be discussing later: the rationalistic paradigm and the naturalistic paradigm.

What is a model?

We should now define a model. Formally, a model is information, data or principles grouped, verbally or graphically (and sometimes mathematically) to represent or describe a certain thing, idea, condition or phenomenon. In less formal language a model is the essence of the learning and thinking of a specialist, stated clearly and briefly. Models are the progeny of paradigms. It is important to remember this parent-child relationship between paradigms and models.

Evaluation models thus can be verbal, tabular or graphic presentations of the principles learned by evaluators. They are the essence of their separate experiences developed in the background of particular paradigms. In other words, they are a set of assumptions, a set of values, a set of preferences and a set of procedures rolled into one.

Finally, evaluators may sometimes talk of evaluation approaches and evaluation strategies. An evaluation approach or an evaluation strategy may be merely a method of beginning or accomplishing an evaluation study. Presumably, when an approach or a strategy, through successive use and testing, becomes both standardized and formalized, it acquires the status of an evaluation model.[2]

SECTION A: Two Basic Paradigms of Evaluation

There are two basic paradigms of evaluation in literacy, development and training:

1. Rationalistic Evaluation (RE), and
2. Naturalistic Evaluation (NE).

This labelling is less than fully satisfactory. The label, RE, has been used as a substitute for logical positivist approaches to evaluation, even though we are acutely aware that it does not fully capture the total set of assumptions of logical positivism. Nor does the label help us to understand that there are now different versions of positivism in use and that a re-conditioned version of logical positivism is emerging. On the other hand, the label, NE, is not too satisfactory either. Just because NE is offered as an alternative to *rationalistic* evaluation, it should not be concluded that naturalistic evaluation is *ir*rational! NE has been used here as a catch-all term for an approach that is at the same time constructivist and collaborative. We shall explain both these terms later. Suffice it to say here that both these evaluation paradigms -- RE and NE -- are "scientific", though they differ in their assumptions about "how the world works" and "what and how we can learn about the world".

The rationalistic paradigm

The rationalistic paradigm is also referred to as the logical-positivist paradigm. It assumes that reality exists "out there" for anyone to see or experience through the senses. In other words, the rationalistic paradigm emphasizes the explicit -- that which is capable of being directly and certainly affirmed. To follow this paradigm is to feel "positive" about the statements one makes about reality; and to depend upon being "logical" in deriving further true statements about reality.

Its three essential features are: *reductionism* -- that parts can be separated from the whole for study without changes in the properties of either; *repeatability* -- that what has been discovered by one should be repeatable by another; and *refutation* -- that what is asserted should be confirmable or refutable. The great hope of the followers of this paradigm is, of course, to generate law-like statements, with universal generalizability.

It is sometimes called the classical paradigm because it has been long in use, follows strict rules and is seen -- not necessarily correctly -- as standard and authoritative.

The rationalistic paradigm follows the methods of hard -- sometimes called restricted -- sciences such as physics, chemistry and engineering. Its methodological ideal is the randomized sample, and controlled experiment. Quasi-experimental designs may be acceptable under some conditions.[3] The rationalistic paradigm demands a clear definition of evaluation objectives and of variables, a sampling plan, structured instrumentation that generates quantitative data, statistical techniques in the analysis of data, and generalizability of results.

The naturalistic paradigm

The naturalistic paradigm assumes that reality does not exist out there for everyone to see and experience in the same way, but that the world is both *found* (as objective reality) and *made* (that is, socially constructed by each individual). Indeed, the most important part of our reality is socially constructed. The evaluator/researcher seeks to find the meanings people carry within themselves. The naturalistic paradigm suggests that human behavior be studied as it naturally occurs, in natural settings, and within its total context. In other words, the naturalistic paradigm is holistic in its orientation, seeking to study reality as a whole, without dividing it artificially into parts and segments to suit the convenience of the evaluator.

The naturalistic paradigm is sometimes referred to as qualitative and phenomenological. This means that, unlike the rationalistic evaluator, the naturalistic evaluator seeks to first describe phenomena and then search for regularities and patterns. The naturalistic evaluator searches for understandings of the specific situation that may later illuminate other somewhat similar situations. The naturalistic evaluator does not search for generalizable laws, but rather for *insights* that can be transferred from one context to another.

In naturalistic inquiry, the methods used are those of the anthropologist and the ethnographer. The evaluator/researcher is himself or herself part of the phenomenon under study -- the evaluator cannot stand in objective isolation "outside of" the reality being studied. The NE design is emergent; it emerges as the evaluator undertakes different steps and follows different procedures in the collection of meaningful data. The samples are purposeful rather than random. The instruments are always unstructured and generate qualitative data. Claims are made in regard to the applicability and fittingness of results rather than to their generalizability.

The rationalistic paradigm versus the naturalistic paradigm

It is not possible to include a more thorough discussion of the two paradigms within the scope of this monograph. In the table on pages 32-33 we have summarized the differences between the two paradigms of evaluation that the reader should examine carefully.

It is now being suggested that many of the claimed strengths of the rationalistic paradigm are merely assertions and no more: some of these assertions are conceptually indefensible, others are impossible to sustain in practice. For example, it is often difficult to select proper criteria for judging the merit or worth of programs; and evaluators using the rationalistic paradigm occasionally end up choosing narrow criteria that would fit their experimental plans. Experimental situations are essentially uncontrolled, irrespective of the claims to the contrary. Randomization in sample selection is quite impossible. Groups under study are frequently self-selected and are systematically different from each other.

Experimental treatments are impossible to standardize across sites. One is never sure about what it is that is being compared. There are limitations to the information which can be produced by the use of this paradigm. Its use produces no information on the process and none on net effects of various interventions. Indeed, much information generated is unusable and is "dead on arrival".

There are, of course, many administrative difficulties involved as well in the utilization of the rationalistic evaluation paradigm. Program needs and rationalistic evaluation needs often pull in different directions. Choice among groups that would receive treatment -- and development resources -- and those that would not, is no easy matter.

The questions of reliability and validity

Too often, discussions of rationalistic evaluation (RE) and naturalistic evaluation (NE) come down to discussion of validity and reliability. Is the data collected objective? Is it valid? Is it reliable? It is now widely understood that to ask these questions in these words is to judge NE by the standards of RE. That is at a practical level unfair, and at conceptual level absurd.

Lincoln and Guba in their recent book[4] have discussed the various aspects of *rigor* as they apply to the two basic paradigms of evaluation. What they are saying is that both the rationalistic and the naturalistic paradigms of evaluation can be scientific and rigorous in their own terms. They suggest that terms appropriate to judge the goodness and dependability of NE are different from those used for RE. See the table below:

RATIONALISTIC AND NATURALISTIC TERMS
APPROPRIATE TO VARIOUS ASPECTS OF RIGOR

Aspect	Rationalistic Term	Naturalistic Term
Truth value	Internal validity	Credibility
Applicability	External validity/ generalizability	Fittingness
Consistency	Reliability	Auditability
Neutrality	Objectivity	Confirmability

What paradigm should evaluators of development, education and training programs choose?

The rationalistic paradigm has had great victories in the hard sciences. It has produced research that has banished diseases from the face of the earth and has put man on the moon.

It was so successful that social scientists (sociologists, psychologists, economists, educators, even anthropologists) wanted to mimic the "scientific" paradigm of the physicist and the chemist. They used the rationalistic paradigm with a vengeance. It made them feel like real scientists! For years and years, the rationalistic method was learned and the rationalistic method was taught in most social science departments of universities.

DIFFERENCES BETWEEN THE TWO BASIC PARADIGMS OF EVALUATION

The Rationalistic Paradigm	The Naturalistic Paradigm
Philosophical roots	
Positivist	Phenomenological
Reductionist	Holistic
Value-free	Value-embedded
Theoretical orientation	
Tests available theory	Uses "grounded" theory
Causal linkages	Linkages of plausibilities
	Mutual simultaneous shaping
Design	
Experimental or quasi-experimental to assure objectivity and validity	Emergent design (or rolling design), assuring resonance without separating knower from known
Setting of evaluation/research	
Laboratory or otherwise controlled	Ecological, in natural context
Sampling	
Random	Purposive, elite, specialized
Size pre-determined	Size determined in use, sample is exhausted when available information is exhausted
Methodological orientation	
Objectives-oriented	Goal-free
Quantitative	Qualitative
	"Thick" description
Instrumentation	
Structured, often interventionist	Unstructured, often unobtrusive
Instruments are sought to be standardized and made independent of evaluator's bias	Evaluator/researcher himself or herself becomes the tool of data collection
Preference for hard data	All knowledge acceptable

DIFFERENCES BETWEEN THE TWO BASIC PARADIGMS OF EVALUATION

The Rationalistic Paradigm	The Naturalistic Paradigm
Data analysis	
Typically statistical	Thematic Content analysis of interviews, documents, and observations
Report	
Statistical-analytical	Descriptive, interpretive Typically a case study
Nature of truth statements	
Generalizable laws	Intuitions about natural covariations of happenings Insights, analogies
Convergent findings leading to prediction	Divergent findings
Single tangible reality	Multiple realities or one negotiated construction of reality in context
Strengths	
Provides good estimates of differences, variations, and correlations when variables can indeed be properly defined and reasonable controls can be established	Responsive, adaptable, holistic emphasis, humanizes evaluation activity
Weaknesses	
In seeking to fit the evaluation questions to acceptable methods and modes of analysis may lead to the choice of trivial and artificial questions and trivial unusable results	The evaluator may get lost in the complexities of real life, may be lacking in interpersonal skills and individual perceptiveness and may end up with meaningless impressionistic statements.

The realization has emerged during the past twenty years or so that the rationalistic paradigm has given social scientists good feelings but not necessarily good findings. We have discovered that too often social life does not fit into the experimental mode. In trying to control variables, we segment human behavior unnaturally and indeed change the very nature of the human behavior being studied. Aggregation of scores and statistical treatments of data may look elegant and impressive but results have been trivial and even misleading. In human behavior, the context is important. We need to study not just behavior but behavior-in-context.

The naturalistic paradigm is more appropriate, most of the time, for the study of human behavior. Once rejected out of hand as subjective and qualitative, it is becoming more and more acceptable. As its methodology becomes clearer and techniques of data analysis are further advanced, the naturalistic paradigm of evaluation will find its rightful place in evaluation methodology.

Should it always be the naturalistic paradigm that should be used by the evaluator of broad impact programs of development, adult literacy education, and development training? The answer is: If not always, most of the time. Lee J. Cronbach's advice seems most useful. Cronbach[5] points out that in the world of education and social change, one can come across two different contexts: the *context of control* where the evaluator can control the social situation to suit evaluation needs (Example: the study of eye movements in looking at a large-size instructional poster); and the *context of accommodation* where the evaluator cannot control the social situation to suit evaluation needs but must accommodate himself or herself to existing realities (Example: the study of frustration and aggression among children on the playground). In the contexts of control, of which there may not be too many, it is all right to use the rationalistic paradigm. In the contexts of accommodation, however, the naturalistic paradigm would make more sense because, by definition, the naturalistic paradigm does not seek to disturb the naturally existing realities.

SECTION B: Models of Evaluation

Against the background of these two general paradigms, many different models of evaluation have been proposed by specialists in the field. But why are there so many evaluation models? Is there not one correct way of doing evaluation?

Earlier in this chapter, we defined a model as the essence of the learning and thinking of a specialist, stated clearly and parsimoniously for communication among professionals and practitioners. There are many different evaluation models, because different specialists have undergone somewhat different experiences in learning and doing evaluation and have used different values and world views in reflecting on their experiences.

Evaluation models are different also because they have emerged within different program settings: within formal education or within out-of-school and nonformal education settings; within mental health settings in an industrialized country or within family life education in the context of a developing country.

Finally, and most importantly, evaluation models are different because evaluation specialists have introduced additional "value" considerations to their initial choices of paradigms. Some evaluation models emphasize a more synoptic view of evaluation, suggesting that we evaluate not only the behavior of our so-called clients but also our own. Some evaluation models suggest the introduction of imagination to our evaluations so that we do not depend only on cold calculation. Some suggest that the unanticipated consequences of program actions may be as important as the intended and the anticipated. Therefore, the model of evaluation should be able to accommodate both the anticipated and the unanticipated consequence. Some suggest that evaluation be conducted as an advocacy and confrontation. Some suggest participative evaluation wherein both the means and ends of evaluation are participatively determined by all concerned -- organizers, professionals, and beneficiaries.

One can see a clear underlying value direction in the development of evaluation models during the last twenty years: (1) there is exclusive or complementary use of naturalistic strategies; and (2) there is a move towards inclusion of the beneficiaries of programs in the design and implementation of evaluations. The key words are holistic and participative.

Some of the evaluation models often referred to in the literature of evaluation will be discussed below. The discussions will be brief. We include in this book a discussion of the evaluation models for two reasons: educational and political. The development practitioner or the literacy worker should have some idea of what different evaluation models exist and what their characteristics are.

This is for his or her education. But there is also a political reason. A literacy worker should be able to justify his or her choices of the model or models; and should be able to hold his or her own against the outside specialist. We should not allow technicians and specialists to browbeat us with the use of unfamiliar names and phrases!

The following models will be briefly discussed below:

1. Tyler's objectives-oriented model
2. Societal experimentation model
3. CIPP model and the EIPOL grid
4. Countenance of evaluation
5. Responsive evaluation
6. Discrepancy evaluation model
7. Transactional evaluation
8. Goal-free evaluation
9. Investigative approaches to evaluation
10. Evaluation as illumination
11. Evaluation as connoisseurship
12. The advocacy model of evaluation
13. Participatory evaluation model, and
14. The situation-specific strategy (3-S) model of evaluation.

1. *Objectives-oriented evaluation*

The objectives-oriented model of evaluation is associated with the name of Ralph Tyler and is perhaps the oldest of the available evaluation models.

Evaluation done under this model seeks to make comparisons of "intended outcomes" with "actual outcomes". In other words, children or adults in a program or project are *tested* to see if *objectives* in regard to acquiring particular ways of thinking, feeling and acting, have been achieved. In practical terms, evaluation becomes equated with testing.

There are some good points in this approach. The approach is focused on outcomes, a concept most easily understood. There is no need to define experimental and control groups which can be disruptive of daily routines in schools and communities and can be quite costly to implement. Measurements reflect clearly stated objectives, hence reliability is not much of a concern. While tests are initially criterion-referenced, they can acquire norm-referenced functions if comparisons are made consistently across sites.

However, there are serious disadvantages to the model. The information generated by tests is too narrow to constitute a sound and comprehensive basis for judging the merit or worth of the total program. The information generated by the model is terminal. It is of little direct use for improving the program.

2. *Societal experimentation model*

This is a model that seeks to experiment with already existing social groups. The society becomes the laboratory.

In the classical experimental mode, the evaluator using this model chooses two groups, one of which receives the experimental treatment, and the other does not. The essential methodological concepts are randomization, control, treatment and comparison.

The proponents of the model had also suggested what were called quasi-experimental designs that are supposed to better fit the realities of the real world. However, serious doubts have recently been voiced against the quasi-experimental designs, by their original proponents themselves.

3. *The Context-Input-Process-Product (CIPP) model and the EIPOL grid*

The *CIPP model* is often associated with the name of Daniel L. Stufflebeam, who has used this model in various evaluation studies.

According to the CIPP model, the sole purpose of evaluation is to produce information useful for decision-makers. Using the systems metaphor and the four parameters of systems (context, input, process and output), the model talks of four types of evaluation to provide information for four types of decision:

1. Context evaluation -- to provide information on the setting, to be able to make planning decisions
2. Input evaluation -- to make programming decisions such as alternative project designs and personnel decisions
3. Process evaluation -- to make decisions related to methodologies and implementation, and
4. Product evaluation -- to evaluate impact and to make recycling decisions

The CIPP model when first proposed combined systems vocabulary with formal research, with its stress on the clarification of evaluation decision needs, structured observation, and the testing tradition of achievement testing in schools. The model adopted the criteria of internal and external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness and efficiency of the evaluative information produced. It was criticized for showing little concern for values. Recent versions of the model have tried to meet some of the criticisms.

In the *EIPOL grid*, Ravindra H. Dave[6] translated the "Output" of systems language (called the "Product" in the CIPP model) into two parts: (i) Learning outcomes and other "intermediary" outcomes of the program, and (ii) Long-term effects of the program on the educational and socio-economic domains. Thus, the four system parameters (Context, Input, Process and Output) became five dimensions or phases of evaluation: Environmental Setting, Inputs, Processes, Immediate Outcomes and Long-term Effects. These five evaluation dimensions are placed against four major phases of a project cycle -- pre-planning, planning, implementation and assimilation -- thereby generating what is called the EIPOL Grid.

4. *The countenance of evaluation*

The countenance of evaluation model is associated with the name of Robert E. Stake. It is so called because Stake talked of two countenances (that is, faces) of evaluation -- description and judgement.

This model was directly related to the evaluation of effects in terms of stated objectives and involves the completion of two data matrices as follows:

	Description Matrix		Judgement Matrix	
	Intents	Observations	Standards	Judgements
Antecedents (Inputs)	x	x	x	x
Transactions (Processes)	x	x	x	x
Outcomes	x	x	x	x

The task of the evaluator is to find data for all the cells in the table above to compare observations to intents; and to make judgements in terms of the standards agreed to among program organizers and evaluators. One should note that in systems vocabulary antecedents are inputs and transactions are processes. The model in implementation has used stratified random samples for collecting special information, combined with the case study approach.

The model has called the attention of evaluators to the need to define standards on the basis of which judgements can be made, though the model itself has left the question of specification of standards unresolved.

5. *Responsive evaluation*

Subsequently, Stake has moved to the concept of Responsive Evaluation -- an evaluation mode that comes closer to transactional and naturalistic evaluations. It is not pre-ordinate (that is, already defined by the evaluator as a specialist) but is responsive to real needs of audiences requesting information. Its focus is not on program intents but actual activities. It is multiple-perspective and uses naturally occurring communication of all those involved. It seeks to collect not only information but also to catch the mood and the mystery of the phenomenon under study. Therefore, it is informal and iterative and emphasizes thick descriptions. As can be seen, it is a very humanistic approach to evaluation.

6. *The discrepancy evaluation model*

The model was proposed by Malcolm Provus, who defined evaluation as the art of describing a discrepancy between expectation and performance of a program.

The basic tenets of the model are standards (S), performance (P), and discrepancy (D). The task is to compare P against S to determine D and thereby to make judgements about the worth or adequacy of an object. The model further suggests that we look for discrepancies in terms of five different aspects of a program:

1. the design of the program
2. its installation
3. the processes of implementation
4. the product, and
5. the cost

On the face of it, the model sounds somewhat rationalistic, but it is not. The model indeed humanizes evaluation and makes it responsive by the manner in which the concepts of standards, performance, and discrepancy are applied. For instance, the evaluator neither sets standards, nor judges the comparisons between standards and performance. The evaluator merely collects performance data and points out the discrepancy. The client must, however, set the standards, though the evaluator helps in the clarification of the design structure of the program and thus in the establishment of appropriate standards. The client, again, should point out what performance information will be most useful for making decisions; and must make judgements about discrepancy.

While recognizing the usefulness of the experimental method in certain cases, the model shows preference for the descriptive methods of history and anthropology and the case study method of sociology and psychiatry. With its relative emphasis on naturalistic methods, it suggests that evaluators work in teams to be able to test individual perceptions of each against the other, and to be able to question the standards being applied to describe discrepancies.

The model claims to provide continuous information to decision-makers on the performance of an on-going program. It also claims to provide information that has a direct one-to-one relationship to decisions actually being made. The resources required for effective application of this model can be considerable in terms of personnel, time and money, however.

7. *Transactional evaluation*

The transactional evaluation model is rooted in transactional psychology, which considers perception and knowing as a transactional process. These transactions deal with concrete individuals, within concrete settings; and the evaluator, as viewer, is always part of the set of transactions. The model is associated with the name of Robert M. Rippey, who has challenged educators and trainers to concentrate on the educational processes -- the program, the classroom, and the school -- rather than on what scores their students and trainees have made.

The focus is on educational accountability -- change-makers are asked to study themselves, their roles, the systems in which they play these roles and the larger systems that surround the systems under change.

The methodologies recommended are informal. In Rippey's own words:[7]

> A comparison with traditional summative and formative evaluations shows that the target of evaluation is different: the subject of evaluation is the system, not the client or the services rendered by the system. The variables relate to the social, psychological and communication aspects of the system, rather than to the manifest objectives. The information is continuously fed back into the system. The evaluator himself is more a part of the operating system. The conventional considerations of reliability, validity and objectivity are less important than those of timeliness, relevance and the observable effects of generating evaluation information. Primarily, evaluation is intended to transform the conflict energy of change into productive activity; to clarify the roles of those persons involved in the program changes, not to produce new knowledge or ascribe causality.

One should note the assumptions in regard to the basic paradigm in use in the transactional evaluation model and the additional value positions introduced in the model. It is indeed a highly value-laden model. It emphasizes relational information and urges sensitivity to the unanticipated consequences. It also implies that evaluation be conducted collectively by the protagonists and designers of a change program and by representatives of those likely to be affected.

8. *Goal-free evaluation*

The idea of goal-free evaluation was introduced by Michael Scriven. He pointed out that in our emphasis on stated goals, our search had become completely focused on intended effects -- effects we wanted to create under accepted program goals. This focus became so exclusive that we often developed a tunnel vision: looking for evidence of intended effects and seeing nothing else.

He suggested that we should look for the real effects of programs, effects that had actually occurred whether intended or unintended. This he thought could be done if we conceived of a goal-free evaluation, independent of objectives stated for the programs. Results from objectives-focused evaluation and goal-free evaluation of a program could then be combined. The use of goal-free evaluation should not suggest evaluation in the so-called "responsive style". In fact Scriven is very keen on summative evaluation and on comparisons that consumers can use. He also suggests that we do more and more personnel evaluation (that is, evaluation of teachers, field workers, etc.) and make people accountable.

9. *Investigative approaches to evaluation*
Jack D. Douglas[8] has analyzed the methods of the investigator or the detective to show how investigative strategies could be used to expose the truth about people in social settings.

The investigative model does not assume a world of cooperation, openness and truthfulness, but one of misinformation, evasions, lies and fronts. He then suggests strategies for grasping an evaluation setting, infiltrating the setting, building friendly and trusting relationships, and then using them in a continuous process of testing out and checking out.

The *modus operandi* model, suggested by Michael Scriven, is also an investigative method for studying cause-effect relationships through sequential testing. This method reconstructs the procedures of the historian, the detective, the anthropologist, and the engineering trouble-shooter. The *modus operandi* model is proposed as a substitute for experimental and quasi-experimental approaches when field situations preclude their use. Essentially, the method involves generating hypothetical chains of cause-effect events and eliminating those that could not possibly have happened. This, of course, is the typical method of the detective.

10. *Evaluation as illumination*
This model was developed in clear rejection of the "agricultural-botany" model of evaluation rooted in the scientific paradigm. It was asserted that groups and communities cannot be randomly assigned to treatments like farms and fields; and human beings cannot be administered treatments like seeds in the ground. In any case, quantitative data generated by the agricultural-botany model provided only partial descriptions of phenomena.

Parlett and Hamilton[9] built this model on two important considerations:

1. Instructional systems, once adopted, become living systems. Living systems do not match their catalog descriptions. Important modifications occur in programs as they move from the drawing board to actual implementation.
2. Programs of training and development cannot be separated from their learning milieu. Actors in the learning milieu and the structures of the milieu become part of the instructional system.

While retaining the use of sampling methods, and structured questionnaires and tests, Parlett and Hamilton drew our attention to the naturalistic methods for description and interpretation. Three stages in the evaluation process are suggested to include: (a) observation of the educational setting; (b) selection of themes through progressive focusing and intensive inquiry; and (c) analysis and explanation.

11. *Evaluation as connoisseurship*

The connoisseurship model of evaluation proposed by Elliot W. Eisner[10] makes a clean break with the scientific paradigm and draws from the aesthetic tradition of the arts. Teaching, Eisner says, is artistry; and schooling is a cultural artifact. Then why not evaluation as connoisseurship? He asserts that indeed a single connoisseur who has spent a lifetime in a field, through the systematic use of perceptual sensitivities, organized past experience and refined insights, can provide evaluations that may be impossible to obtain in any other way.

Eisner suggests two interrelated concepts: (1) educational connoisseurship and (2) educational criticism to perform the tasks of educational evaluation. Educational connoisseurship is the means through which the shape of the context and the configurations within it can be reorganized so that intelligent decisions about the context can be made. Educational criticism is the art of disclosure through description, interpretation and evaluation.

The methodology of connoisseurship and criticism is by no means soft-headed or romantic, and certainly can be systematic and rigorous. Educational critics can learn to look for the pervasive qualities of education in the classrooms and training settings; and can learn to look for the meanings of hidden cues. Questions of reliability and validity must be handled through *structural corroboration* (mutual validation of one bit of data by the rest, the whole being supported by the bits that constitute it); and through *referential adequacy* (the existence of a relationship between what the educational critic says and the subject matter of his or her critique). Generalizations are also possible in the sense that educational criticism will lead to more refined processes of perception in subsequent settings; and will create in the evaluator's mind new anticipations.

Reports of educational criticisms have a family resemblance to case studies, but case studies of educational criticism are different in the sense that criticism itself is an art form. As a critical disclosure, an educational criticism report creates a living image, communicating to its readers a visceral understanding of the educational realities.

12. *The advocacy model of evaluation*

The advocacy model is also called the adversary evaluation model or the judicial evaluation model. As the name suggests, this model uses quasi-judicial procedures in the conduct of evaluation. Typically, two groups of people both for and against a program are allowed to advocate their opposite positions before an educational jury in terms of issues generated and selected for the trial. Evidentiary rules and procedures are established and cross-examination is permitted. It is an educational trial by jury.

Proponents of the model cite several advantages of the model. It enables evaluators to develop and use explicit procedures for generating and assessing alternative program strategies; provides a record of decision-making for later accountability; accommodates not just data but also perceptions, opinions, biases and speculations; and can involve a variety of stakeholders in the trial.

On the other hand, there are those who have found serious faults with the model. The model unnaturally dichotomizes positions as "for" and "against" a program. In real life, of course, there are not two but many sides to the same issue. The model changes evaluation into a competitive event. Since groups are assigned sides by the flip of the coin, there is often a mismatch of "lawyers" and a lack of conviction in defending positions. Judges and those who sit on juries vary in their abilities.

In view of the many negative aspects of the model and the huge expense involved in mounting a trial, the "court case" format has been changed into what are called "clarification hearings". Juries have been eliminated, though some sort of a panel may still be used. Expert witnesses may be called for positions both pro and con. There may be some cross-examination. The issues are thus clarified, but decisions about preferences and modifications are left to the listeners.

13. *Participatory evaluation model*

The name of Paulo Freire, the Brazilian educator and the author of *Pedagogy of the Oppressed* (New York: Herder and Herder, 1972) is often associated with participatory evaluation and research. A considerable amount of work has been done in this area during the last ten years by evaluators spread all over the world. Participatory research networks have been established, participatory research and evaluation studies have been conducted and their results published.

Participatory research or evaluation is not a scientific endeavor of the professionals, but an in-depth, existential review of an experience done by all concerned, together, in collaboration. The learner becomes an evaluator and the evaluator becomes a learner. Evaluation goals, ends, standards and tools are decided upon participatively. Each contributes personal data and collects the data that has to be obtained. Analysis of data is collectively undertaken. Judgements are also rendered collectively.

In an address to the Institute of Adult Education, University of Dar es Salaam, back in 1972, Paulo Freire presented the possible steps in such a participative methodology:.

1. The evaluation (or research) team should acquaint itself with all previous research and evaluation -- no matter what methods were used in that previous evaluation or research.
2. The team should delimit the area of action geographically -- even though, culturally speaking, there are no frontiers.
3. The team should identify official and popular institutions in the area selected and go to talk to the leaders within those institutions.
4. The evaluation team should tell these leaders, in all honesty, that they have come to discuss the possibility of *all* people in that community holding discussions and working together.
5. If the leaders agree, the evaluation team should hold meetings not only with the leaders of various institutions but also with the people who are involved in some way with those institutions.
6. The evaluation team should discuss with the community arrangements for meetings wherein groups of, say, thirty people could come together on a daily or weekly basis for discussions. Such meetings might involve almost all the inhabitants of a community and last for several weeks. The important thing would be to obtain a perception of the whole community.
7. Sociologists, psychologists, educators and linguists should, at this stage, join the research or evaluation team and visit each group. Records of discussions should be made at each

meeting. People should be urged to speak if they are silent, but otherwise the role of the evaluation team should be no more than advisory. One of the members of the community should chair such meetings.

8. Justice, education, government, industry and many other topics may be discussed; but all in terms of the people and in the context of concrete realities.
9. When the smaller groups think they have exhausted the topics for discussion, each one should put its findings on paper and then they should all meet in a general session. The reporters at such sessions should be the people themselves; not the specialists on the team. The workers should become intellectuals. There should be collective discussion of each group report.
10. The evaluation team should now make a critical study of the people's discourse. This study should be interdisciplinary. The various levels at which people perceive reality must be determined and their many implications should be worked out. These implications must be studied in the presence of the people, not by social scientists on their own.
11. The evaluation team together with the people should now draft a proposal for subsequent action. The programme itself should not be worked out *for* the people but *with* the people.

It should be clear from the preceding that participative evaluation is not distinguishable from need assessment or community awareness. The distinction between evaluation and instruction as well gets lost in participative evaluation. Participative evaluation provides participants with further opportunities to raise their consciousness and consolidate their sense of power and self-worth.

As indicated above, considerable work has since been done in participatory evaluation by the International Council for Adult Education, which has established a participatory evaluation network all over the world. There do not seem to have been any significant methodological departures, however. Their methods have more or less retained the spirit of Paulo Freire's list of steps given above.[11]

14. *Situation-specific strategy (3-S) model of evaluation*

Before presenting our 3-S evaluation model, let us remind readers that it is useless to look for *the* model of evaluation, or for one correct way of evaluating literacy, training or development. As

Cronbach has reminded us, one model may fit the "context of control" and another the "context of accommodation". A literacy worker might often be using more than one of the above models, within the context of a single evaluation study.

Another important point to remember is that models are not usable as formulas. Models are to *think with*. They would seldom give you unchangeable sets of procedures, step by step. When they do, they would probably mislead.

The 3-S model to be discussed below is an empty set that should help us select the right model or the right mix of models and approaches to be used in an evaluation program or an evaluation study. The conceptual essence of the 3-S model is this: Do not start with an evaluation model, begin with the evaluation problem. Analyze the evaluation problem into sub-problems; think how the problem or parts of the problem might unfold over time; and, finally, think of the milieu in which evaluation will be conducted.

Different parts of the evaluation problem will most likely require different evaluation models and approaches. You may need both a survey and an in-depth case study. You may require achievement testing of learners as well as content analysis of documents.

The exigencies of time may demand pulse-taking through quick appraisals, even though, ideally, a more systematic evaluation would have been better. Finally, the evaluator may be working in a situation where there are no calculators or colleagues who can help with the analysis of large bodies of numerical data; where there are no copying machines or stencil duplicators; or where there is no duplicating paper for producing the required instruments. The 3-S model helps us think about what strategies to choose in specific real-life situations, about how to do "the second best" when the very best is not possible.

Elsewhere,[12] we have listed the following steps in the implementation of the 3-S evaluation model:

1. Articulating the means-ends relationships in the change program to be evaluated
2. Generating profiles of information needs and evaluation issues
3. Developing a situation-specific evaluation agenda
4. Choosing appropriate and realistic methodologies and techniques

The 3-S model permeates the evaluation planning and evaluation management approach presented in this book.

Things to do or think about

1. Of the two basic paradigms discussed in this chapter, which is likely to generate more useful information on your development, literacy or development training program? Or, do you have to use a mix of both?

2. Of the models described in this chapter, which model or models do you personally consider most useful in your work at this particular time? What more would you like to know about the model to put it to use?

3. Can you find evaluation studies already completed that fit *neatly* under one or the other model described in this chapter?

Notes

1. Kuhn, T.S. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.

2. In our discussion in this chapter, the term "theory" has not been defined or explained. This is so because the literature on evaluation talks often of evaluation models and seldom of evaluation theory. Let us say briefly that, in terms of the conceptual status, theory falls between the paradigm and the model. In its best sense, theory is a deductively connected set of laws and empirical generalizations. A model is often a schematic diagram that connects theory with practice.

3. Reference is being made here to the work of Donald T. Campbell and Julian C. Stanley, *Experimental and quasi-experimental designs for researchers*. Chicago: Rand McNally, 1966.

4. Lincoln, Yvonna S. and Guba, Egon G. *Naturalistic inquiry*. Beverly Hills, CA: Sage, 1985.

5. Cronbach, Lee J. et al. *Towards reform of program evaluation.* San Francisco: Jossey-Bass, 1980.

6. Dave, Ravindra H. "A Built-in System of Evaluation for Reform Projects and Programmes in Education." *International Review of Education,* Vol. 26 No. 4, pp. 475-482, 1980.

7. Rippey, Robert M., ed. *Studies in transactional evaluation.* Berkeley, CA: McCutchan, 1973, pp. 3-4.

8. Douglas, Jack D. *Investigative social research.* Beverly Hills, CA: Sage, 1976.

9. Parlett, M. and Hamilton, D. "Evaluation as illumination: A new approach to the study of innovatory programs." Occasional Paper, No. 9. Edinburgh: Center for Research in the Educational Sciences, University of Edinburgh, 1972. See also Richards, Howard, *The Evaluation of Cultural Action. An Evaluative Study of the Parents and Children Program (PPH).* London: Macmillan, in association with the International Development Research Centre, 1985.

10. Eisner, Elliot W. *Educational imagination: The design and evaluation of school programs.* New York, NY: Macmillan, 1979.

11. "Participatory research: Developments and issues." A special issue of *Convergence,* Vol. XIV, No. 3, 1981.

12. Bhola, H.S. *Evaluating functional literacy.* Amersham, Bucks, U.K.: Hulton Educational Publications Ltd., 1979. Pages 25-33.

Part II

Evaluation Planning and Management

In this Part II of the monograph, the interrelated processes of "evaluation planning" and "evaluation management" are discussed. An evaluation planning approach (EPA) and an evaluation management approach (EMA) are explained and demonstrated.[1] It should be noted that these two processes of evaluation planning and evaluation management are conducted at the program (or institutional) level. Thus, they must precede the process of "design" and "implementation" of individual evaluation studies. This Part is divided into the following chapters:

1. Evaluation Planning, and
2. Evaluation Implementation and Management.

CHAPTER 3

EVALUATION PLANNING

The concept of "evaluation planning" is relatively new to the literature of evaluation. To plan is to choose. Evaluation planning is to choose from among the many possible evaluation questions. To generate a set of significant questions, system thinking is necessary. The evaluation planning approach demonstrated in this chapter suggests that all interlinked systems -- the literacy system, the community or the performance system within which literacy will be utilized, and, finally the surrounding social system -- be described in terms of the four system parameters, that is, context, input, process and output. Questions should then be raised about what parameters need illuminating and, consequently, what information should be generated in order to clarify what is unclear. The ideal set of information needs should then be subjected to the criteria of desirability and feasibility. The shaken-down list of information needs of decision-makers should then form the evaluation agenda for a particular program.

"Planning" and "plans" are today familiar words in most parts of the world. Typically, a plan is a set of intentions or arrangements worked out in advance; a method, scheme or design for the attainment of some objective in the immediate or the distant future. In everyday life, planning is the more or less *intuitive* process of developing such a plan.

Professional evaluation planning

In the professional life of an evaluator, the essential meanings of plans and planning remain the same. However, for an evaluation plan to be so called, it must be more than merely intuitive. It must result from the planning process which has been deliberate, systematic, informed, and rooted in reality. These criteria are more likely to be met if the evaluator implements a process of planning suggested in the following:

1. Evaluation must be conceptualized as a response to the information needs of decision-makers. Further, this evaluative response must be organized to be both *systemic* and *systematic*. It should be systemic in the sense that it involves system thinking. The evaluator must see the evaluation exercise as linked with the literacy program system, the community or the performance system in which literacy skills will be utilized, and the surrounding social system, all at the same time. The various systems should be described in systemic-dynamic terms using system parameters -- input, context, process and output.
2. The evaluative response, again, must be systematic in the sense that the choice of evaluation questions is not arbitrary. It should not allow the evaluator to get stuck with the very first evaluation problem that is somehow thrown up. It should demand a look at the totality of information needs -- *first, and every time* -- before particular choices of questions are made and particular data collection strategies are chosen.
3. Dynamic descriptions would involve questions such as this: what inputs, through what processes, under what contexts, lead to what outputs? The evaluation planner, with the assistance of decision-makers themselves, should then list the various information needs of decision-makers arising from these dynamic descriptions, separating the urgent and the feasible from information that is merely "nice to have".

What is a system?

In the above listing of steps, we have repeatedly used the word system. At this point, it is necessary to introduce a formal definition of a system. A system is an orderly arrangement or combination of interrelated and interdependent parts or elements emerging into a whole. A family is a system. A cooperative is a system. A literacy program is a system -- a techno-social system, we might add. We live, breathe, work, vote, play, and pray within social systems of various kinds.

Systems and sub-systems

Systems may have sub-systems within them. Sub-systems may, in turn, be composed of sub-subsystems. On the other hand, systems may be part of larger supra-systems and mega-systems. It is important to remember that boundaries of systems and sub-systems are not God-given. They are boundaries that we assign to systems simply because we have found those boundaries convenient for both understanding of and intervention into systems.

System thinking

System thinking is the mental habit of looking at things as a *whole*. It is "holistic thinking." It is the type of thinking that enables us to think of multiple processes happening together in "at-once-ness" and helps us avoid the pitfalls of linear thinking. System thinking is to learn to look at various entities and individuals as connected together into network of relationships even as they appear separate, and isolated.

System descriptions

A most important advantage of system thinking is that all systems, whatever their nature, size, or complexity can be described using the same set of four parameters. The four parameters are input, process, output and context. The "process", as a system parameter, lays bare the dynamics of a system. The "input" as a system parameter tells us what the system is living on. We can ask ourselves the question: What variation might be possible in inputs and processes to get different and more preferred "outputs" in a particular "context"? In some cases, the "context" itself may be manipulable. The point to note is that a description of a system in terms of these four parameters can be called a *description in design terms* or a *dynamic description*. Since our ultimate objective in evaluating functional literacy or post-literacy is to intervene in the teaching and development processes to improve them, such descriptions are most useful.

System descriptions of "Literacy for Development": Three interlinked systems and sub-systems

By way of demonstrating the process of developing what we have called "descriptions in design terms", we take the example of evaluation planning of a "literacy for development system". We will show that to engage in evaluation planning in this case, the evaluator cannot avoid dealing with a literacy program system, the community or performance system in which literacy will be practised, and with the overall socio-economic system. Figure 1 on the next page shows these relationships graphically.

The literacy system receives relevant inputs which are subjected to particular processes in the specific social/organizational context of the literacy program. Some desired (and some unanticipated) outputs result. The "XYZ" in the graphic are outputs which did not come from the literacy system itself but were added to the literacy system outputs from outside to become inputs for the community/performance system.

Of course, the literacy program system should have been so designed that it was in perfect interface with the community (or the performance system) within which literacy skills will be utilized. It is amazing how often this obvious requirement is neglected by planners. Within the community/performance system, once again, the inputs are subjected to processes in a particular context to produce outputs from this particular system. The community/performance system, again, will produce outputs that will be both anticipated and unanticipated. Some of these will be what were desired, others will not be desired. These outputs will be supplemented with other outputs "PQR" from elsewhere within the social system and will become inputs into the overall dynamics of the social system.

Describing systems in dynamic terms

There are, of course, many different ways in which one could describe the societal system undergoing development, a particular community/performance system, or a literacy system serving adult learners. As indicated earlier, the four system parameters (input, process, context, and output) provide the best system descriptions. The table on pages 58-59 demonstrates the point and should be examined. The tabulation is self-explanatory. The listing of inputs, processes, outputs and many layers of contexts for the three interlocking systems described in the table is not necessarily complete.

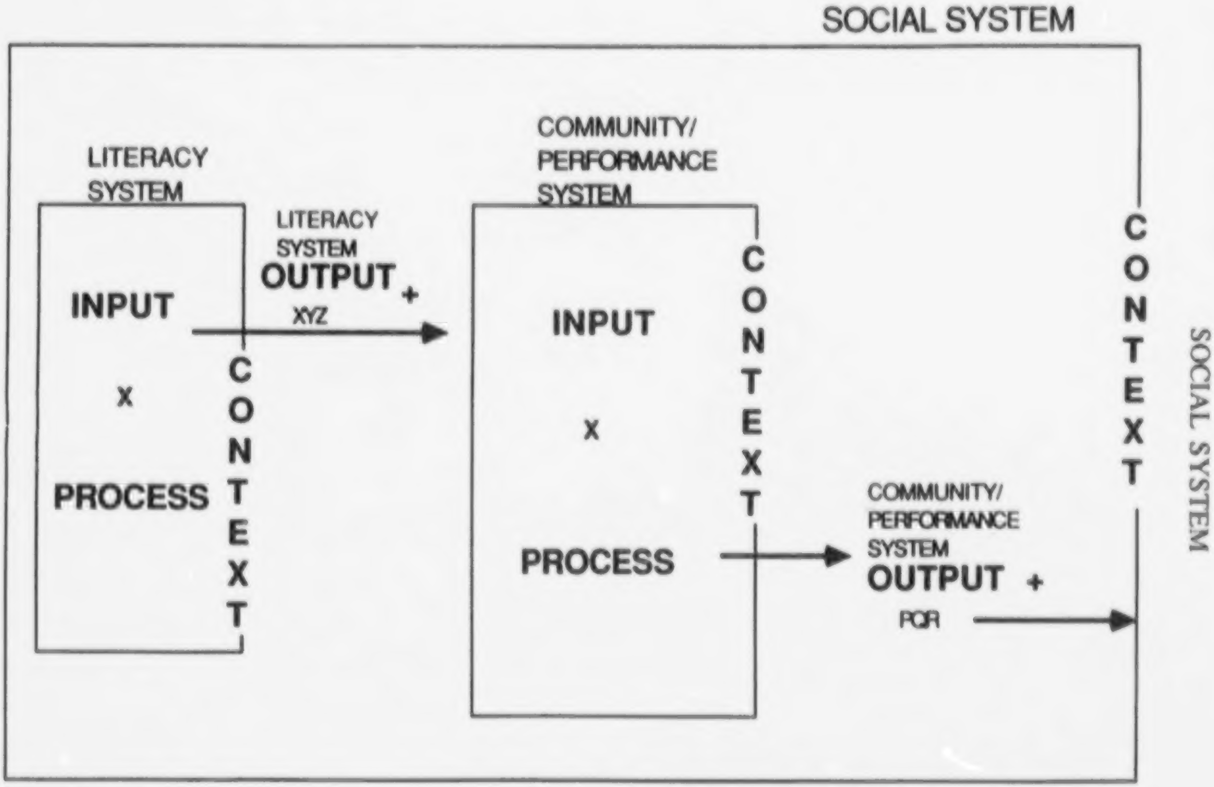

SOCIAL SYSTEM
LITERACY SYSTEM
INPUT
x
PROCESS
CONTEXT
LITERACY SYSTEM
OUTPUT +
XYZ
COMMUNITY/ PERFORMANCE SYSTEM
INPUT
x
PROCESS
CONTEXT
COMMUNITY/ PERFORMANCE SYSTEM
OUTPUT +
PQR
CONTEXT
SOCIAL SYSTEM

Figure 1: A Model for Evaluation Planning for "Literacy for Development" Initiatives

	LITERACY SYSTEM	COMMUNITY/PERFORMANCE SYSTEM	SOCIAL SYSTEM
INPUTS	Learners -- male, female Facilitators -- teachers, extension workers, political educators Methods and materials Technological inputs Local infrastructures	Inputs from the functional literacy system Post-literacy and continuing education facilities Rural/urban libraries Extension services Vocational training in factories Packages of credit and information	Input from the Community/ Performance system Ideology, political will Policy initiatives in culture, development, media Technology/ infrastructures
PROCESSES	Educational (formal, informal) Extension Awareness-raising Mobilization Social reorganization Coordination Management	Educational extension Management training Second socialization Organization Institution building Mobilization Cultural renewal Staff training	Modernization Democratization
CONTEXTS	Group climate Social organization-- age sets, peers, etc. Local culture Community politics-- factionalism, casteism Sexism, ageism Community's learning environment	Community politics Social organization Agricultural estates Cooperatives Factory organization Educational/cultural infrastructures	National International

	LITERACY SYSTEM	COMMUNITY/PERFORMANCE SYSTEM	SOCIAL SYSTEM
OUTPUTS	Functionally literate individuals Politically aware individuals Tested materials and methods Experienced facilitators Effective local systems Better learning environment	Literates as users of literacy, making more effective transactions with all aspects of environment -- economic, social, political, physical Dynamized infrastructures Fewer accidents School enrollment of children Cultural renewal	Modern society Democratic society

Generalized functional literacy is assumed with its three components -- (i) literacy skills, (ii) functionality and (iii) awareness. Both (a) rural and (b) urban contexts are reflected.

Generating information needs

We can now move to the third and last step of evaluation planning: Developing a set of information needs for the decision-makers. Looking at the table above, we can go through the various entries in the cells of the table and consistently ask the same set of questions:

1. Do decision-makers have sufficient information on the various *elements* appearing or embedded in various cells?
2. If not, what are their information needs?
3. Is the information needed, possible or feasible to collect? (Some of the variables may not even be manipulable by the decision-makers. In that case, more precise information may not help much.)

On the basis of such a table, we may be able to generate a list such as that on pages 60-61 about the literacy program system for use in generating information needs and then specific evaluation agendas.

PARAMETERS /	
Variables	Variations/Options
INPUTS	
Teachers	Particular educational levels and particular social class. Extent of field work experience. Level of commitment to development work. Teaching competence and teaching experience. Direct appointment versus secondment from a parent department. Continuity versus turnover. Workloads of teachers.
Learners	Educational background. Social class and value orientations. Commitment to development. Motivation to learn.
Teaching materials/ facilities	Teaching materials -- quantity, diversity, quality. Materials and facilities. Indigenous versus imported instructional materials. Instructional and duplication equipment. Characteristics of learning sites.
PROCESSES	
Instructional/ Informational	Conceptualization of teaching as knowledge transfer, skills training, behavior modification, socialization, etc. Integrated versus discipline-oriented curriculum development (i.e., instructional organization). Teaching and learning styles. Substantive knowledge versus process emphasis. Presence versus absence of curriculum validation through needs assessment. Availability or nonavailability of counseling and guidance services.
Organizational/ Structural	Organizational health status. Organizational capacity rating.
Distributive/ Maintenance-related	Quality of administrative support. Coordination with extension staff and services.

PARAMETERS / Variables	Variations/Options
CONTEXTS	
Organizational	Organizational culture. Institutional relationships (horizontal and vertical) with other organizations.
Environmental	Surroundings (Closeness to a bar versus a "retreat" situation). General social climate in the country.
OUTPUTS	Literate adults at various literacy levels. Trained development workers with various competences. Emergent role identities. Differential experiences of trainers. Quality of radio programs.

A list of this type must now make the evaluator confront "What is" with "What can be". The evaluator should now look back critically on his or her day-to-day experiences within the program and try to articulate clearly the problems which were there but perhaps were hard to get hold of. The evaluator should also look at the existing teaching-learning system and the program system critically and think of the higher returns that could be obtained by making some changes.

In all these cases, the evaluator should be able to state some information needs: We have the problem "X", but we do not have the information "Y". Or, if we had the information "Y", we could take the promising step "X" with confidence. Two important points must be mentioned here:

1. *A distinction should be made between evaluation problems and administrative problems.* To administer is to direct and superintend the execution, or conduct of a program. If administrators, for reasons of incompetence or for lack of responsibility fail to direct and superintend a program, the problem is one of administration, not of evaluation. Evaluation can only assist administration by providing needed feedback data and by testing

various program assumptions. It is not a substitute for administrative decision-making.

2. *Evaluation may require conceptual analysis or collection of framework data, and not field data.* There will sometimes be evaluation questions which will have to be answered through conceptual and operational analysis rather than by going to the field to collect data. Is the participatory method recommended in a training program actually employed in the training protocols? Is the integrated curriculum concept actually embedded into the training plans, training materials, and training delivery and schedules? These questions require analytical answers and not necessarily collection of data. Again, policy documents, the nation's five-year economic plans and census data may have to be used to build a framework for the evaluation of an aspect of the literacy campaign, program or project.

From evaluation questions to evaluation agendas

The evaluation questions generated in the step above may all be interesting and promising but it may not be possible to answer all of these in the particular context of a literacy program and within the resources available. In such a situation, a particular evaluation agenda must be followed within a particular time period.

The following criteria might be useful in the choice of evaluation questions for inclusion in the evaluation agenda:

1. Availability of options for intervention
2. Significance of the evaluation question, and
3. Feasibility of implementing the evaluation study.

1. *Availability of options*

All of the variables entering a literacy situation may not be under the control of the decision-maker. In other words, the literacy trainer may not be able to change the values of the variables in any significant way. If such is the case and program variables are "immutable", it is no use evaluating them because they do not offer options for re-design.

2. *Significance of the evaluation question*
If a variable does offer an option for intervention, it will make sense to evaluate it, if in a relative sense, it offers a significant option. The significance has to be in terms of the effectiveness or efficiency of results. In either case, the returns from the evaluation effort should be worth the effort.

3. *Feasibility in regard to available resources*
The evaluation question chosen and the evaluation design that is necessary for conducting the evaluation should be within the capacity of the literacy project, program or campaign. A reasonable amount of resources should be available for evaluation to avoid unnecessary frustrations.

The concept of *evaluability* often referred to in evaluation literature these days is addressed to concerns similar to those discussed above.

Things to do or think about

1. Using the INPUT-PROCESS-CONTEXT-OUTPUT table discussed above, analyze your own literacy program in terms of various elements.

2. What do you need to know about the most significant demands made upon literacy workers and change agents by the "performance system" in your case? Re-state these information needs in the form of a set of evaluation questions.

3. What are some of the important facts of socio-economic and political life in your environment that must be reflected in an evaluation agenda?

Note

1. See also Bhola, H.S., *Evaluation planning, evaluation management and utilization of evaluation results within adult literacy campaigns, programs, and projects.* Bonn: German Foundation for International Development, 1981. [ERIC Document No. 221 759]

CHAPTER 4

EVALUATION IMPLEMENTATION AND MANAGEMENT

Evaluation implementation and management as discussed in this chapter must cover three separate but interrelated processes. First and foremost, the best mix of effective and efficient strategies of information gathering must be developed. These strategies must be within the resources of the program or institution concerned and must together generate data that is timely, credible and confirmable. This strategic mix, we suggest, will have to involve the methodological triangle of MIS (Management Information System), NE (Naturalistic Evaluation) and, perhaps, RE (Rationalistic Evaluation). Second, the management function must establish an appropriate institutional context for evaluators to play their role: that is, the evaluator role must be interfaced with the programmer role to reduce role conflict to the minimum. Third, the management function must assure actual utilization of evaluation results by making the program administration a learning culture that uses evaluation data as a matter of habit.

Research management (or administration), over the past two decades or more, has become a speciality of sorts. However, there has not been a transfer of the management concern to the evaluation area. At best, evaluation management is seen as a matter of organizing for data gathering in the field without breakdowns. Evaluation management problems are much more extensive, however.

It is important to be reminded here of the fact that we are not yet talking of evaluation *design* -- that is, the technical design of evaluation studies. The design questions will arise later within the context of each individual study. In discussing the evaluation management approach, we are raising and answering questions prior to the design question.

The meaning of management

Management is a fuzzy word. It is often equated with administration. Management has recently come to be seen as a more comprehensive process that includes planning, organizing, implementing, and controlling the work of others.

Our concept of evaluation management includes a mix of *professional* and *organizational* decisions. The professional decisions must involve choices among and between the strategies of information-gathering so that the information gathered meets the criteria of effectiveness, efficiency and timeliness. There are also important organizational decisions involved. These relate to the invention of evaluation roles and designations; and their placement within the program system in such a way that there is a minimum of role conflict within the organization.

The methodological triangle of evaluation

On the basis of our experience with evaluation training and evaluation practice during the last fifteen years in the Third World, and particularly in Kenya[1], Botswana[2], Malawi, and Zimbabwe, we are able to assert that at the institutional (or program) level, the strategy of information gathering must consist of the methodological triangle of evaluation as shown in Figure 2.

MIS: Management information system

We shall discuss an MIS (Management Information System) more fully in Part III below. In the present context, we need only to discuss it in its barest details.

A Management Information System (MIS), as the name suggests, is an information system that assists in the effective and efficient management of an institution, program or project. Important information about a development program or a literacy project is organized into an information system so that it can be systematically stored and easily retrieved for use by decision-makers. Typically, the information is such as is generated by the program or project in the very process of its implementation. While computer hardware and software may be used in the development of an MIS, neither is necessary. Paper and pencil management information systems are possible if they use well-designed registers, forms and tables filled on a daily or periodical basis.

We are firmly of the belief that the establishment of an MIS should always be a necessary part of all evaluation management approaches. Start-up costs, especially of paper and pencil MIS's, are low. Returns on the costs of establishing MIS's are high.

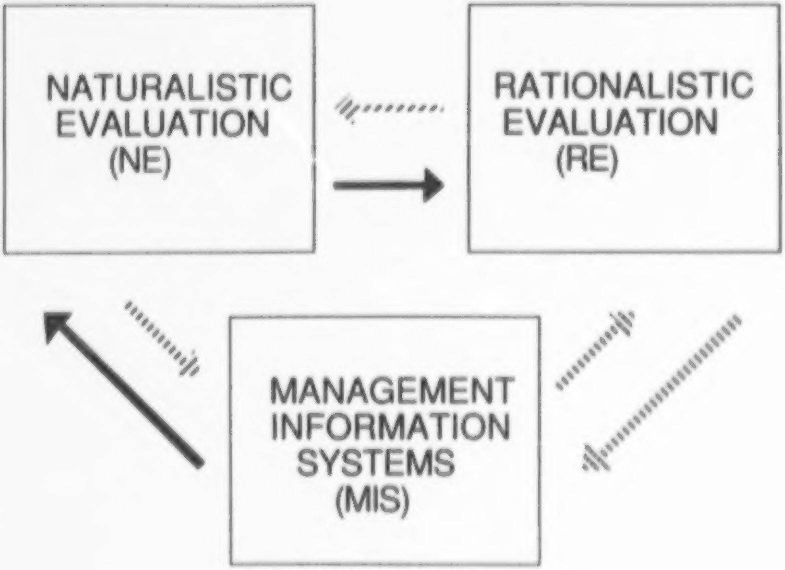

NATURALISTIC EVALUATION (NE)
RATIONALISTIC EVALUATION (RE)
MANAGEMENT INFORMATION SYSTEMS (MIS)

Figure 2: A Model for Evaluation Planning and Management in the Context of Program Implementation and Policy Assessment

It has been found that MIS data are the data most widely and most frequently used by decision-makers making day-to-day decisions about a program. How many people are there in a program? How many are men and how many are women? Where do they live and work? How did they score on the literacy test? What quantities of fertilizers did they buy? How much did they produce per acre? A well-designed MIS can answer all these questions on a regular basis and can help decision-makers make good implementation decisions.

We look at the MIS as a necessary component of the methodological triangle of evaluation management. An MIS will, by itself, serve many important information needs of decision-makers. In addition, an MIS will support both NE (Naturalistic Evaluation) and RE (Rationalistic Evaluation) activities. Contrary to the often-held but naive belief, NE does not mean all qualities and no quantities! While the naturalistic evaluator does seek to obtain "meaningful constructions" rather than "meaningless counts", it does not mean that the naturalistic evaluator does not know how to count. NE does make use of numerical data; and, of course, of descriptive statistics when these are useful in developing meanings about multiple realities. The point to be made here is that the MIS would often support NE: first, by encouraging the evaluator to go beyond numbers and look for the meanings about change held by the various stakeholders; and, second, by becoming the anchor for such meanings.

NE: Naturalistic evaluation

In Chapter 2 we presented a rather detailed description of the naturalistic mode of inquiry. To recollect, it was pointed out that the philosophical roots of naturalistic inquiry were "phenomenological" and "systemic". NE used "grounded theory" developed within the "social context" of the inquiry itself. Its samples were "purposive". It did not pretend to become objective by taking the evaluator out of the process of inquiry but used the "evaluator as an instrument" of observation. NE developed "thick descriptions", which were examined for recurrent "themes" and were "interpreted" to develop meanings for participants in their own realities. The statements about reality were not universal laws but statements about "multiple realities" as different stakeholders experienced them.

The world of a development trainer or of a development worker teaching functionaries new skills of education and extension or promoting dissemination and incorporation of new knowledge, skills and attitudes is, in Lee Cronbach's words, a "context of accommodation" as contrasted with a "context of control", which is typically available only in the laboratory. Understandably, therefore, we consider NE as an important component of the Evaluation Management Approach suggested here. As was asserted above, the MIS data is the most widely and the most frequently used in decision-making by development planners and trainers. We will now suggest that qualitative statements about the success or failure of a program made by those who have come in direct contact with the program are the second most widely and frequently used "information" for decision-making.

We firmly believe that in the all-pervasive context of accommodation in the world of development, literacy and training, NE can be used to develop "qualitative" statements about the life of a program or about the effectiveness of training that are much more dependable and credible than common-sense statements made by various constituents and stakeholders. These qualitative statements, when read with numerical data from the MIS, will give decision-makers most useful information for improving an on-going program or project.

In a later part of the monograph (Part IV), we shall discuss the process and techniques of conducting naturalistic inquiry.

RE: Rationalistic evaluation

In Chapter 2 we discussed the concept of rationalistic inquiry in some detail. Essential features of RE were described. It was considered "positivist" and "reductionist". It sought to test hypotheses generated from a theory. Its sampling was meant to be random and it used experimental or quasi- experimental designs. Its instruments were structured and data analysis was typically statistical. Its aim was to develop assertions that were objective and had the validity of universal laws.

We are now realizing that educators (as well as other social scientists) cannot often use RE. This is so because the world of development and change does not belong to the context of control. Ours is a messy world of multiple variables, of entities that are not reducible to single aspects to suit our purpose. The assumptions on which the methodologies and techniques of RE are built do not in fact *always* hold in the real world.

Then why is RE part of the triangle of the Evaluation Management Approach we have offered? We include it for two reasons: one professional, the other political. The professional reason is that while conditions for RE (that is, the context of control) may not exist often, they will exist sometimes. In that case, RE methodology may be used to make statements on some aspects of the program. The political reason is that NE as a methodology has still to win full legitimacy with evaluators working all over the world. Decision-makers in the Third World who themselves may have been trained in the classical methodologies of RE are not going to suddenly accept NE methods just because we say so. The love affair with RE is going to take time to break up. Then, evaluators are going to take their time to fall in love with this new girl, NE, that somehow seems to make sense but is rather complex, untidy and uncertain about things!

Readers should note the relationships between MIS and RE and NE and RE as shown in Figure 2. The contributions of MIS to RE can be easily surmised. There will be times when data already in the MIS could be fitted into the evaluation design developed in the RE mode. At other times, new data may have to be collected which could then become part of the MIS already in place. But MIS and RE are not one and the same thing. The two are significantly different. The primary and essential use of MIS data is to profile the size, scope and surface structure of a project, program or campaign. RE studies reduce the world to dimensions, characteristics and variables, and seek to make normative and thereby generalizable statements about that reality.

The relationship between studies in the NE and RE modes may not be so obvious. Quite often it can be (and perhaps should be) that NE studies generate some questions that can be studied in the context of control of RE studies. At other times, an RE study may demand that a more meaningful statement be made in addition to the general sort of comparison between groups or between before and after behaviors of the same group.

In Part V of the monograph, we shall describe the details of RE as a mode of inquiry.

Role relationships between evaluators and programmers

Too many evaluation efforts are doomed to failure because of a defective conception of the evaluation subsystem in relation to the total program system. Unfortunately, some evaluators have too exclusive a conception of their roles. They think that they are the only ones functioning as evaluators and everybody else is there to answer their questions, to do their bidding and feel and act respectfully. They fail to realize that in an institution or program that is engaged in internal evaluation, there will be fewer full-time evaluation officers (FTEO's) within the system, than full-time program officers (FTPO's). These FTPO's will be playing significant though only part-time evaluation roles. In terms of the time spent, their inputs into evaluation work will indeed be many times more than the man-hour inputs of full-time evaluators.

As part of the Evaluation Management Approach, all FTPO's must be made to understand and accept their evaluation roles. These roles should be made clear and concrete. The relationship between the FTEO's and FTPO's must be made clear as well. It should be understood that the evaluators are there not in the inspectorial role but in the technical role of enabling the FTPO's to be able to conduct evaluations of their own work.

Working with outside consultants

In some Third World countries, in the near future, it is likely that national evaluators will be working with outside experts who are assigned to a program on a short- or long-term basis. It is impossible to write a script for a relationship between the local evaluators and the outside consultants in any great detail. Some general ideas should, however, be laid down.

The initiative in evaluation planning and evaluation management, and later in the implementation of evaluation, should always remain with the local people. The outside consultant may be asked to collaborate in, even to guide, the process of evaluation planning, but should not be asked to provide full-fledged evaluation plans. It is for the local evaluators and decision-makers to decide the "What" question. The outside consultant can provide reinforcement as an expert.

The same is true of the "How" questions. The outside consultant should not be simply ordered to produce, on his own, the best possible design which can then be followed by the local evaluators. Here, again, the initial ideas must come from the local evaluators and the design should then be worked upon together with the outside consultant.

The best way, therefore, to use a consultant is to use him or her as a formal trainer and an informal socializer. The consultant should never be used as one more staff member added to the program team.

Creating a learning institutional culture

There is considerable discussion in evaluation literature on the topic of utilization of evaluation results.

It has been suggested that the utilization of evaluation is an exception rather than the rule. Too often evaluation questions studied by the evaluation team are meaningless for the decision-maker. Oftener the results of evaluation are available after the fact and too late for use. The political climate and the issues may have changed so that the results of evaluation are best forgotten and put on the shelf to collect dust.

It has been suggested that perhaps we expect too much from "utilization" and that quite often we do not see how evaluation studies may indeed have influenced decision-makers. Decision-makers may, that is, have in fact responded to informal presentations of results as they emerged during data collection, and to interim reports as they were written and made available. In many cases, the direct trail of influence may not be available, and yet a study may indeed have been most influential.

In the case of internal evaluation, utilization will be more likely even though it may not always be obvious. Indeed, in internal evaluation, people at various levels of the system start responding to information without the benefit of a formal decision by an authority. This informal utilization can be further enhanced by higher level authorities by involving people at various program levels in the design and implementation of the evaluation process and by concrete declarations of intent on the part of the institution or the program to make informed professional decisions.

Creating a learning culture within communities

This process of participation and norm setting can indeed be extended to leadership within communities. By modelling behavior of the appropriate kind, the community leadership can be taught to let politics bow to information, at least to make information an important part of the politics of development at the local level.

Things to do or think about

1. Do you have an MIS in your project? If not, do some rudiments exist in the form of registers and forms and periodical reports, etc.?

2. Have any evaluation studies been undertaken within your program or project? Can you separate them as instances of NE or RE?

3. Do you think that field visits from the headquarters to the field can be converted into some sort of one-man naturalistic evaluations? How?

4. What are some of the evaluation topics related to your program that fit better in the RE mode?

5. What kind of evaluation (NE or RE) would higher level decision-makers in your country prefer? Why?

Notes

1. Bhola, H.S. *Action Training Model (ATM) -- An innovative approach to training literacy workers*. N.S. 128. Notes and Comments. Paris: Unesco, Unit for Cooperation with UNICEF and WFP, March 1983. Also Bhola, H.S., "Training evaluators in the Third World: Implementation of the Action Training Model (ATM) in Kenya." *Evaluation and Program Planning*, Vol. 12, pp. 249-258, 1989.

2. Bhola, H.S., "Building a built-in evaluation system: A case in point." Paper presented to the Evaluation Network (now American Evaluation Association), San Francisco, October 1984. [ERIC Document No. ED 256 779.]

Part III

MANAGEMENT INFORMATION SYSTEM (MIS)

The evaluation management approach discussed above in Part II, Chapter 4, consisted of what we have called the methodological triangle of evaluation: MIS (Management Information System), NE (Naturalistic Evaluation) and RE (Rationalistic Evaluation). It should be restated that the MIS is easily the most important component of the three-pronged strategy of information gathering. It can be further suggested that in the allocation of evaluation resources the first priority should go to the development of an MIS, howsoever rudimentary such an MIS may be. Our discussion of the MIS in this Part of the handbook will be divided into the following chapters:

5. MIS -- Theory, Questions and Design

6. Writing a Proposal for Developing an MIS

7. The Process at a Glance: Tools and Techniques of Implementing an MIS:
    - Section A: Concept Analysis
    - Section B: Writing Indicators
    - Section C: Making Tests of Achievement
    - Section D: Testing Attitudes, Observing Actions and Results
    - Section E: Data Analysis, and

8. Writing Periodical and Special Reports Based on MIS Data.

CHAPTER 5

MIS -- THEORY, QUESTIONS AND DESIGN

The MIS may be seen to be conceptually rooted in two interrelated ideas: some information is better than no information in decision-making; and action programs, typically, generate useful information in the very process of their implementation that can, in turn, be used for decision-making. The essential principles for the design of an MIS are rather simple: The basic dynamics of the program that an MIS will serve should be structured as a system; the variables clustered under each of the four system parameters (inputs, pro-cesses, outputs, and contexts) should be identified; indicators should be developed for those variables that cannot be directly seen; sources of data should be identified; a paper and pencil (or computerized) storage and retrieval system should be developed; and a routine about periodicity of data inputs and reporting to decision-makers should be put in place.

Most of us, most of the time, tend to think in terms of black *or* white and miss the greys that are made of black *and* white. We seem to find clear and direct opposites much easier to handle than complex situations where opposites commingle and coexist.

Evaluators and researchers are not immune to this tendency to think in terms of pure opposites. Evaluators, too, continue to make watertight divisions between those who quantify and those who use qualitative methods. They seem to think as if those who use qualitative methods do not know how to count beyond ten! On the other hand, those who quantify are dismissed as "number-crunchers" -- as if they are interested merely in numbers and never in meanings of things.

It is important to remember that "quantity-quality" is not a dichotomy but a complementarity and does not, therefore, by itself, set NE against RE. Indeed, the use of the "qualitative methodology" does not ensure "naturalistic evaluation". RE can first collect qualitative data, and then quantify it for processing! In other words, both NE and RE can use qualitative approaches to data collection. NE does use qualitative approaches more frequently than RE, but there is much more to NE than the use of qualitative methods of data collection. What distinguishes NE from RE is the fact that a different set of assumptions about reality and our knowledge about this reality are involved.

Similarly, naturalistic evaluators are not allergic to numbers. Many among us continue to think that naturalistic evaluators perhaps have no interest in numbers. And since a Management Information System (MIS) is a system for the storage and retrieval of numerical data, therefore, naturalistic evaluators have no interest in an MIS. Nothing could be farther from the truth.

In Part II, Chapter 4, we have suggested that an MIS is the necessary component of any evaluation management approach; and that an MIS would support all other evaluation efforts, irrespective of whether they are undertaken in the naturalistic or the rationalistic mode. It does not hurt to be "informed", whether one is a naturalistic or a rationalistic evaluator.

What is an MIS?

An MIS (Management Information System) is, as the name suggests, a system of information for management.[1] A system is designed which can be used for storage of data. These data are typically numerical -- though an MIS will often be complemented by files containing policy and planning documents, instructional materials, photographs, films and videos. The data are typically collected at fixed intervals of time. To meet both the regular and emergent needs of decision-makers, these data are retrieved to develop information that is useful in decision-making.[2] The focus on decision-making is so important that the label DSS (Decision Support System)[3] is taking the place of the old name, MIS. By combining databases with artificial intelligence, it is now common to speak of Expert Systems (ES's) that assist decision-makers more effectively than old MIS's.

MIS theory and methodology

Theoretical and methodological issues of evaluation and research can be stated in the form of two interrelated questions: What is the nature of reality? and How should we go about making knowledge-full or informative assertions about that reality?

At the deeper theoretical and methodological level, MIS's are rooted in positivism as discussed in Part I, Chapter 2 and to be summed up later in Part V, Chapter 13. At the surface level, the theory and methodology of MIS's can be presented as follows immediately below.

Decisions, of course, often are and will continue to be taken intuitively, without the benefit of information. There will sometimes be instances when information available is too little, or not very dependable. The idea of an MIS is rooted in the simple concept that *informed* decisions are likely to be better than uninformed decisions; and that every effort should be made to collect, and store for later retrieval, information that is dependable and sufficient for day-to-day management decisions.

A related concept, seldom made explicit, may be that *numbers* provide useful content to qualitative statements. Here, for example, is one kind of statement: "It was perhaps the largest football crowd this high school had ever seen." Add to it the following: "All the 15,875 available seats in the stadium were filled." We can see how the numbers support the emotion of the qualitative statement. The MIS data, obviously, will support evaluations in the rationalistic mode, but their support to naturalistic evaluation should not be neglected or underestimated.

Finally, the conceptual underpinning of an MIS is provided by the fact that all programs generate data (both quantitative and qualitative) in the very process of their implementation. Particularly, quantitative data are the easiest to collect and most bureaucracies do collect such data as a matter of course, both for internal and external accountability. With some self-conscious and systematic effort, these data can be fitted into a well-functioning and useful Management Information System. We should hasten to add that Management Information Systems can be inexpensive paper-and-pencil systems, though a low-cost micro-computer and appropriate software should not be considered to be out of the reach of most programs in the Third World. After all, one can today buy ten or more micro-computers for the price of a Landrover.

Materials to supplement an MIS

As indicated above a good MIS must be supplemented by printed and pictorial materials such as national policy and planning documents, training materials, related printed and audio-visual materials used in the program, newspaper clippings and so on. Such contextual and supplementary materials will be necessary for converting numerical data into information usable for making decisions.

Typical questions that an MIS can answer

An MIS typically includes numerical data, but numerical data need not merely represent quantities. Qualities can be assigned numbers and thereby made part of an MIS. A comprehensive MIS can include data on inputs, processes, outputs and context. It can deal with social units from individuals to groups, institutions, and communities.

By providing data according to various time series, MIS can give us information on the structures of programs, on levels and pace of achievement of learners, on effectiveness of curricula and on program impact on communities.

In addition to before and after and time series information, MIS data can easily be used for understanding correlations, for example, between literacy and numeracy; and for understanding differences of achievement, for example between male and female achievements along different indicators, or those resulting from different teaching methods and different post-literacy materials.

The list of questions suggested below is by no means exhaustive, but does indicate the likely usefulness of a well-designed properly functioning MIS:

Questions about program size and structure

1. What is the number of learners in the program by (i) region, (ii) ethnic origin, (iii) gender, (iv) age-set, (v) occupation?
2. What is the number of groups of learners by (i) gender (all male, all female, mixed), (ii) rural/urban location, (iii) institutional setting or location, (iv) linkage with functionality or awareness, (v) months/years of incorporation?
3. What is the pattern of stability and change in program participation over time, and in relation to expressed motivations?

4. What is the present level of achievement (and/or retention) by learners of (i) literacy and numeracy, (ii) functionality, (iii) awareness? Who got to what level, in what time and stayed there for how long?
5. What is the number of teachers in the program by (i) gender, (ii) age, (iii) insider/outsider status in relation to the community, (iv) educational qualification, (v) primary occupation?

Comparisons and differences over time

1. What are the patterns of differences in regard to (i) learner motivations, (ii) stay in the program, (iii) regularity of participation, (iv) achievement?
2. What are the differences in achievement in relation to (i) primary occupations of teachers, (ii) teaching methods, (iii) language of literacy, (iv) overall curricula?

Correlations between entities

1. What is the correlation between (i) achievement in literacy and numeracy, (ii) numeracy and functionality, (iii) teacher qualification and learner achievement in general?

Questions of impact

1. What have been the differences (positive or negative) over time in indicators of (i) quality of life within communities, (ii) political participation, (iii) community health, (iv) preservation of environment and cultural assets, (v) creation of a literate environment?

The questions above related to comparison, correlation and impact will also appear under RE later in the book. The difference between the two approaches (MIS and RE) is that MIS seeks to make statements on the size, scope and surface structure in order to present a profile of the program, while RE conducts studies using random samples and selected assumptions for design and statistical analysis, in order to be able to make normative, generalizable statements.

Designing an MIS

Design essentially is the practical, task-specific aspect of the theory and methodology of something. In MIS the design issues can be handled as a set of operations as follows:

1. Describe the dynamic structure of the action system first in common-sense terms of actors, means and ends; and then translate the description in terms of the four system parameters: inputs, processes, outputs and contexts.
2. Visualize your program system to be fully functioning under ideal conditions and list all the possible variables, again, under the four system parameters: inputs, processes, outputs and contexts, for comparisons and review.
3. Putting the ideal and real side by side, select those variables which will give the MIS a completeness and integrity; and on which information must be collected or generated for making day-to-day decisions.
4. Define, elaborate and analyze concepts where necessary, and develop indicators for those concepts and variables which are not available for direct observation.
5. Identify sources of data on selected indicators that will be put into the MIS.
6. Develop a paper-and-pencil system of registers, forms and tabulations (or obtain appropriate computer hardware and software).
7. Develop a complementary filing system to include documents in print, pictures, film or tape.
8. Establish intervals for data inputs and retrieval, and patterns of data flow for reporting to decision-makers.

Using the examples of paper-and-pencil MIS's for functional literacy programs in Botswana and Malawi,[4] we provide brief demonstrations for each of the steps for the design of MIS's listed above.

1. *Describe the action system in system terms*
Let us recollect our discussion on systems and system descriptions in Part II, Chapter 3: Evaluation Planning. The conceptual structure of the program in question must be fully understood. Begin with the system as it really is in your context and develop a full

programmatic description of the program: The functional literacy action system involves (1) literacy teachers, collaborating with (2) extension workers, and supported by (3) teacher trainers and (4) program administrators, using (5) teaching materials and (6) other instructional processes, to teach (7) adult learners, typically, in (8) the group setting of a literacy class, (9) new knowledge, attitudes and skills that learners could apply (10) at home, (11) in the farm or factory and (12) communities for (13) individual and national development. This programmatic description should then be cast in the terms of four system parameters:

INPUTS
Learners
Literacy teachers
Extension workers
Teacher trainers
Program administrators
Teaching materials

PROCESSES
Teaching and related instructional processes

OUTPUTS
New knowledge, attitudes and skills
Better homes, farms, factories
Individual and national development

CONTEXTS
Literacy groups

2. *Describe the variables of an ideal-type system under the four system parameters*

The variables of an "ideal-type" functional literacy program under the four system parameters should then be listed. For instance, we could have under:

INPUTS
Learners
Teachers
Extension workers
Local leaders
Trainers
Administrators

Instructional materials
Facilities
Related resources

PROCESSES
Instructional
Distributional
Organizational

OUTPUTS AND OUTCOMES
Newly literate adults
Better trained teachers
Better trained trainers of teachers
More effective community leaders

Better instructional and training materials
Better developed infrastructures

Innovation adoption
Higher productivity

CONTEXTS
Instructional contexts
Organizational contexts
Political contexts
Cultural contexts

Such a description of an ideal-type system will enable the designer of an MIS to review the actual system in action both (i) for *purposes of programmatic change* as well as (ii) for the *purposes of anticipating information needs*. The evaluator will thus be able to design an MIS which has completeness and integrity. Naturally, we cannot and need not collect data on all of the variables listed above for our MIS. Selection of important variables will be necessary.

3. *Selecting important variables to put into the MIS*
Data collection and storage cost money. Too much data may in fact clutter the system and make data retrieval and use less likely. This means that the designer of an MIS must understand the various policy issues as well as the various program directions and possibilities available to decision-makers. These understandings will serve as criteria for selection of the variables that should be put into the MIS. In the functional literacy programs of Botswana and Malawi, decisions were made to collect data only on the following:

ADULT PARTICIPANTS IN LITERACY CLASSES

Name
Age
Sex
Previous education
Occupation

Date of joining the literacy class
Expectations from participation in the class
Level of participation
Reasons for absences

Level of skills attainment at various intervals of time
Achievement in other development knowledge and skills
Attitudinal changes

Uses of literacy
Practice of new attitudes

LITERACY CLASS/GROUP

Location
Accessibility
Lighting, ventilation, seating

Date of establishment
Name of supervisor
Name of teacher
Does the teacher live in the same community?

Learning cycle(s), and pattern of expected progress
Potential participants
Ratio of participation (Participants in class/Potential participants)
Number of times a week the class is held
Days of the week the class is held
Duration of class

Number of times class was not held during the period of report

Learner status: Active, Dropouts, Repeaters
Attendance ratios (Actual attendances/Possible attendances)

Teaching quality
Collaborations with other extension workers

Achievement patterns
Applications, uses of learning

FAMILIES OF PARTICIPANTS IN THE PROGRAM

School enrollment of children
Participation in literacy classes and other development groups
Use of family-oriented innovations: family planning, nutrition, hygiene, other
Purchase of durable goods
Community participation as leader/member

FARMS WORKED BY PARTICIPANTS

Size
Crops and related land use
Adoption of innovations
Income

COMMUNITIES IN WHICH PARTICIPANTS LIVE

Name of village/district/province
Distance from the main road
Population (M, F)
Age distributions

Occupations
Additional income-generating activities

Number of farms
Farm sizes
Crops and other land use

Number and types of extension agents
Levels of innovation adoption

Literacy/Illiteracy rate
Number of literacy classes/centers
Coverage ratios for M, F

Availability of elementary school
School enrollment (M, F)
Percentage not enrolled (M,F)

Radios
Newspaper readers
Religious institutions
Secular institutions

Presence and functioning of village development committee(s)

Gross economic output

TEACHERS CONDUCTING LITERACY CLASSES

Name
Age
Sex

Marital status
Children

Family occupation
If part-time teacher, other occupation of self

Residence, with postal address

Date of employment

Education
Literacy training: Year, duration, achievement scores
Other development training

Radio listening
Reading habits: books, newspapers.

Some additional information was to be collected about the production, distribution and storage of instructional materials. It must be stated that data collection plans were substantially reduced after some experience in implementation of the MIS.

4. *Concept analysis, developing indicators and codes*

Some of the variables we have listed above are a matter of observing, asking and recording, such as: age, sex, attendance, occupation, farm size, etc., etc. But many other variables are not available for direct observation. Indeed some are not even definitionally clear.

Those that are not clear will have to go through the process of concept analysis. We shall have to define what we mean by concepts such as "dropout", "adoption of innovation", and "community participation," etc.

Concept analysis may not be enough. Even after some concepts have been analyzed, their component parts may not be concrete enough to be seen and observed. The concept of "community participation" could be concept analyzed into its components: economic participation, social participation, cultural participation and political participation. But then we will have to find indicators for all of these various aspects. Take the example of political participation. We can see whether people attend political meetings or go to vote at election time. Both of these could be used as indicators of political participation.

In some cases, indicators may have to be given values not through observing and asking, but by "tests". We may decide that a score on a literacy test will be an indicator of a person's achievement in literacy. That means that a test will have to be constructed and administered. Similarly, we may develop indicators for motivation to learn or for national integration, and then "scales" may have to be constructed to give values to indicators of change in attitudes. "Observations" may have to be used for collecting data on indicators of utilization of literacy skills or adoption of innovations.

In some cases, "codes" will have to be invented for storing information in the MIS. For example, instead of marking adult learners simply as P-Present or A-Absent, one could use various codes: P-Present, S-Sick, T-Travelling, F-Attending Funeral, U-Unknown, etc. This will make a lot of information available to MIS simply by using a well-designed code.

5. *Establishing sources of data*

The sources of data may be people, groups, institutions, and communities: learners, trainers, community leaders; women's clubs, discussion groups; health clinics, rehabilitation centers; and villages. Sometimes these may be physical entities such as homes, fields, shops, wells, storage bins, etc. As part of this step one should ask questions about who will supply data and with what level of aggregation. What will be the coverage? What groups will be covered and which ones not? Questions about available data sets should also be raised. Will it be possible to merge data sets? What is the possible general quality of data to become available from different sources?

6. *Developing a paper-and-pencil/computer-based system for storage and retrieval of data*

A perfectly workable paper-and-pencil MIS can be established. All information needs will have to be fitted into various registers, application forms, grade books, diaries, logbooks, and periodical reports. Some of the information will have to be repeated in more than one form. Computer-based MIS's are no more beyond the reach of most developing countries. Simple but useful programs for MIS's can be developed for use in micro-computers.

7. *Developing complementary files for materials in print or on film*
Due attention should be given to the development of files of print, graphic, film and video materials. These files will often provide both contextual and illustrative materials for use in reports to decision-makers, developed from data in the MIS.

8. *Establishing time series for data inputs, data flows and utilization of data*
Time series must be established and these must be, on the one hand, realistic -- monthly data collections may be impossible -- and, on the other hand, timely -- six monthly data inputs may be too infrequent to be of any use to decision-makers.

Patterns of data flow must be established as well. Who will collect what data, do what with it, and send it to whom, in what form, and when?

Problems of data utilization are the most important. There must be reports written for use by decision-makers and the reports must take the form in which they are most easily usable. Too often, administrators appropriate to themselves the right to interpret data sent upwards by the various functionaries in the system. They then issue orders downwards on the use of this information. In the meanwhile, functionaries wait even though they are knowledgeable about what is happening and what should be done. My suggestion on this point is this: Never send data upwards without first having collated, interpreted and used them to understand and improve your part of the program. It is also most important to establish patterns for the weeding of "dead information" from the MIS.

Concluding remarks

As has been often asserted in its behalf, the MIS does assist in the understanding of the processes embedded in social action; it enables on-going policy analysis and planning; makes it possible for program functionaries to study the impact of their own programs; and it increases public accountability.

Things to do or think about

1. Develop a programmatic description of the program you are working in. Reorganize the variables involved in terms of the four systems parameters. What questions arise in your mind from doing this in regard to the goodness of the program's conceptual structure?

2. What are the various forms, tables, registers, and reports already in use in your program? Do they all together constitute a good enough MIS? What would need to be done to develop an effective MIS on the basis of what is already available?

3. Find out if a development program or a development department in the country already has a micro-computer available for use. What use are they making of the micro-computer? If they operate an MIS, ask them to demonstrate to you how the software (the computer program) works.

Notes

1. There are numerous books published on the subject of the theory and design of Management Information Systems. By way of an example see: Chacko, George, *Management information systems*. Oxford: Pergamon, 1979.

2. The relationship between data and information implied in this statement should be noted. Data do not speak for themselves. Data are used to develop information needed by decision-makers.

3. See Alter, Steven L., *Decision support systems: Current practice and continuing challenges*. Reading, MA: Addison-Wesley, 1980.

4. The MIS for the functional literacy project is described in Government of Malawi, *Functional literacy programme: Guidelines for monitoring the programme*. Lilongwe: National Centre for Literacy and Adult Education, Ministry of Community Services, Government of Malawi, December 1983. Revised 1989.

CHAPTER 6

WRITING A PROPOSAL FOR DEVELOPING A MANAGEMENT INFORMATION SYSTEM (MIS)

A literacy project, program or campaign could generate a lot of information, both quantitative and qualitative, in the very process of its implementation. By introducing appropriate tables, forms, instruments and reporting requirements, this information could be collected and stored for later use. The possibilities are almost limitless. But collection, storage, retrieval and processing of information costs money. Good managers, therefore, like to collect and store only that information which is necessary and sufficient in the context of their program objectives and resource constraints. To do this, a good evaluation planner and manager must begin with a proposal for developing an MIS to meet the special needs of a particular literacy initiative.

In Chapter 3 of the book, while dealing with evaluation planning, we pointed out how an ideal literacy system -- campaign, program or project -- will have a whole array of information needs to be met partly by an MIS and partly by specially designed NE, or RE type studies. Later, in Chapter 5, the information needs best served by an MIS were listed. Even these relatively narrow information needs to be served by an MIS can be overwhelming.

An MIS proposal should typically include the following items:

1. *The background and program context*

The proposal for an MIS should begin with a brief statement of a country's development policy and the role assigned to literacy and post-literacy in the promotion of such development. The ideology, objectives, and strategy of the literacy campaign, program or project should then be described. The development components (such as fertility, income generation, awareness etc.) with which literacy will have direct interactive relationships should be indicated.

2. *Justification for an MIS*

Establishment of an MIS will consume precious resources. The proposal, therefore, should justify why an MIS should be developed. How will the MIS help in program management, in improvement of teaching and in the study of impact? Will the MIS save resources by improving the internal efficiency and the effectiveness of the program system? Finally, are the data which it is proposed to put into the system available elsewhere from other sources?

3. *Information agenda*

The "information agenda" for the literacy initiative in question should be finalized. Literacy information systems can have their own special foci. A particular MIS may focus on learners and not teachers. In learning outputs, an MIS may focus on literacy, but not on functionality and awareness. Finally, not much may be put into the MIS about impact on communities, which may be left to be assessed through special evaluation studies.

4. *The MIS structure*

The overall structure of the MIS must be conceptualized as part of the proposal for an MIS. The structure must be defined in terms of two dimensions: hierarchical (levels) and chronological (periodicity). For each level, questions of social units to be covered, indicators to be used, forms to be used for collection and collation of data, information flow, and information utilization should be clarified as follows:

LEVEL A (Field Work Level)
- a. Social units to be reflected
- b. Social indicators to be reflected
- c. Forms, tables, instruments to be used for collecting, consolidating and storing data
- d. Information flow up and down the system
- e. Utilization of information for decision-making

LEVEL B (First Supervisor Level)
- a. Information to be received from below for consolidation
- b. Information to be generated at own level
- c. Information flow up and down the system
- d. Utilization of information for decision-making

LEVEL C (District Level)
a. Consolidating information received from below
b. Adding information generated at own level
c. Information flow up and down the system
d. Utilization of information for decision-making

LEVEL D (Regional/Zonal/Provincial Level)
Same as under C above

LEVEL E (National Level)
Same as under C above

5. *Periodicity in data collection and reporting*
Some data will become available every day in the life of a program. Other data may have to be specially collected through tests and questionnaires. The periodicity of such data collection and reporting should be clearly established.

6. *Overall design of forms, tables, and reporting formats*
Design of forms, tables, and reporting formats is more than a matter of drawing some lines on paper. Paper sizes should be selected that are easily and cheaply available and are easy to store in standard filing drawers and cabinets. Rows and columns should anticipate the space requirements for various responses expected to be inserted therein.

7. *Labelling, coding, and numbering*
Forms should be suitably labelled, ensuring uniformity of labels and nomenclatures. Do not use "learners" in one place and "adults" in another. Do not use "groups" in one place and "classes" in another. Of course, both words can be used, calling literacy classes "classes" and income generating groups "groups". Coding should be according to a system that is congruent with the structure of the MIS. One code, with differing numbers on various forms, may suggest a cluster of forms that go together by levels or units of study. Codes should be such that assist memorization and are less likely to be confused with other codes.

8. *Printing and distribution*
The proposal should anticipate problems in working with the printer. It is, of course, important to work closely with the printer. Useful differentiations can be made in the form or table by use of lines and screens. Again, if forms are not properly and efficiently distributed all efforts will prove useless.

9. *Training for implementation, and utilization*
Installation of an MIS is more than a matter of distributing forms. Actors in the total program system, from literacy teachers who will fill class registers to provincial officers and those in the national headquarters, must be trained to contribute data to the system and to use it for their decision-making purposes. The MIS proposal should anticipate training needs and suggest training plans. These training sessions need not be too long and can be conducted by peripatetic teams. The good news is that in training personnel for using the MIS system, we will at the same time be training them in planning, management, and program design.

10. *Location of records*
The movement of data from one level to another need not be accompanied by the movement of records -- actual registers, log books, forms, and other instruments. Depending upon the material, in each separate case, decisions should be taken as to what records should be located where.

In the remainder of this chapter, we have taken the example of an MIS system actually developed for the National Literacy Program in Malawi and talk of the decisions made during the process of its development.

Example of a Proposal for an MIS

The proposal outlined on pages 96-104 about the design, installation and utilization of an MIS for the National Adult Literacy Program (NALP) of Malawi was developed within the context of a national workshop, "Training Workshop on Management Information and Testing System in the National Adult Literacy Program," held in Zomba, Malawi from May 21 to June 3, 1989.

THE ZOMBA TRAINING WORKSHOP

Two aspects of this exercise in MIS proposal development are worth noting:

First, the workshop was not developing a new proposal but was engaged in a *revision* of an MIS first designed in 1983 and then reviewed in 1985. The MIS needed to serve the needs of a literacy program that had gone from a pilot to a national program. In the meantime, some problems with the MIS had become known and needed fixing. This makes the Malawi case even more useful than it would have been otherwise.

Second, the Zomba workshop was a workshop with a *dual purpose*. It would develop a revised proposal for the design, installation and utilization of an MIS, and it would train functionaries at various levels of the NALP to install, implement and utilize the MIS.

This second aspect of the workshop is worth noting. To ensure the actualization of the "dual purpose" of the workshop, participants were selected so as to represent functionaries from all the levels of responsibility within the NALP -- literacy teachers, supervisors, district officers, provincial officers and those from the national headquarters. On the one hand, this heterogeneous group of participants gave the workshop their special perspectives on the needs, problems and possibilities of an MIS. On the other hand, they received training in the whys and wherefores of the system and were ready to conduct training of their colleagues on return home.

On the basis of our experiences in developing MIS's for various literacy campaigns, programs and projects in Africa and elsewhere, we strongly recommend this strategy for the design of MIS's in literacy promotion systems elsewhere. Of course, a workshop such as this that designs the MIS can also serve as the first in the cycle of training courses for all the functionaries within the system.

i. *Description of the NALP*

Even though all the functionaries at the workshop (with one or two exceptions) were from within the National Adult Literacy Program, it was necessary to begin the proposal development process by discussing the policy objectives, administrative structures, implementation strategies and general information needs of the NALP. A short session on these topics was offered to the participants and additional relevant documentation was made available to the participants. This was time well spent since it enabled participants at various levels to realize their particular obligations to collect information in a timely fashion, to be aware of the information flow up and down the system, and to learn about the utilization of MIS information in decision-making.

ii. *Justification for an MIS in the NALP*

In the context of the example being used here, a justification was not needed for establishing an MIS, though justifications were needed for a second thorough revision of the MIS after four years of use. The administrative and technical problems actually experienced during the use of the MIS provided the justification for the expansion and revision of the MIS. These problems were as follows:

Technical problems

- The existing MIS was incomplete in regard to the post-literacy program and to some other aspects such as the listeners' groups under the radio program.
- Programs of non-governmental partner institutions were also not suitably reflected in the present MIS.
- Data on instructors included in the MIS were insufficient.
- The supervisory staff did not always have the know-how about how to handle the MIS; and many could not decipher or analyze the errors in the reports they received or compiled.
- Some of the registers were too complicated and detailed to use and needed simplification.
- Some information was missed from forms in use, and some information was collected more than once.
- There were still some definitional problems with some of the terms used in the current MIS, and these needed standardization and clarification.
- In some cases one form was confused with another. It was not always clear who would fill a form, who would receive it, and where it would be kept.
- More information was collected than the national HQ could use.
- The MIS component involving Learner Testing was in comparative neglect. For example, there were no figures available about the number of adults declared literate.
- Test items actually used in the present tests for learners had not been organized in an appropriate hierarchy. Therefore, the available tests did not help discriminate between literacy levels.
- There was a lack of know-how on the part of supervisors about administering tests. The lack of know-how about marking tests and preparing results was even more serious.
- Time needed for administering the testing system had been grossly underestimated.

- There was a need to design equivalent tests so that the same test would not have to be administered every time.
- Confidentiality of the process of testing and of results had to be protected.

Administrative problems

- There were problems with the transportation needed to go into the field to collect data, and problems with postal services to send the data up and down the system.
- The unavailability of forms and instruments and lack of stationery presented another set of serious problems.
- Many officials simply did not complete the forms and tables or did not do so in time.

iii. *The information agenda*

In the context of the current proposal, the information agenda for the MIS did not have to be prepared afresh. The information agenda established in 1983 and updated in 1985 remained more or less intact. The focus was still on literacy skills. Data on functionality and awareness as well as on impact on communities were left to be collected through specially designed evaluation studies.

The MIS proposal, therefore, included the following items:

- All forms were to be revised and made consistent with one another.
- All forms were to be given new and mutually consistent code numbers, indicating how the various forms are interlinked, beginning from the village level and rising to the national level.
- New post-literacy activities were to be reflected in the appropriate forms.
- A new form on Radio Listeners' Groups and another called "Guidelines for Supervisors" were to be designed.
- The levels at which data would be aggregated were to be identified so as to include: supervisors' level, district level, regional level, and national level. (This would mean that all information to the Headquarters would be routed through the Regional Offices and not sent directly from the Districts to the Headquarters. A new form for part one of the Regional Co-ordinators' Reports, therefore, had to be designed.)
- More qualitative information would be collected as part one of each monthly report, to support numerical data collected as part two of each monthly report.

- Partner agencies would report through the District Co-ordinator to regional and national level.
- A complete set of forms and instruments would be accompanied by a set of clearly written guidelines on how to use those forms and instruments.

iv. *The MIS structure by levels*

Five roles/levels were identified within the MIS for Malawi, as follows:

Instructor
Supervisor
District Co-ordinator
Regional Community Development Officer
National Center HQ Staff

v. *Periodicity*

Questions about periodicity were not all settled within the context of the proposal, but were left for later administrative decisions. Data about learner participation would, of course, be generated every day. But decisions would have to be made about testing schedules. Again, district reports would probably be written every month, but regional and national reports could be written on a quarterly basis.

vi. *Overall forms and instruments*

For reasons of space all the forms and instruments drafted for the Malawi MIS cannot be reproduced here. Interested readers are referred to Josef Müller and Anja Dietrich, (eds.), *Dossier of evaluation instruments for literacy programmes*, Bonn: German Foundation for International Development, 1989.

A list of all the instruments proposed and designed for the Malawi MIS is included on pages 100-102.

NEW TITLES AND CODE NUMBERS FOR THE INSTRUMENTS OF THE 1989 VERSION

Earlier titles	New titles	Earlier Code	New Code
I. *Instructor's Forms*			
The FLP Class Profile and Progress cum Attendance Record	Class Profile with - Enrollment Sheet - Class Attendance Register - Visits by Extension Workers - Reports on Literacy Meetings	CP	I - CP
First Month Supplementary Report	Instructor's Initial Class Report	MR - IA	I - ICR
Monthly Report	Instructor's Monthly Report	MR - IB	I - MR
II. *Supervisor's Forms*			
Village Profile	Village Profile	-	S - VP
Instructor's Profile	Instructor's Profile	MR - IP	S - IP
First Month Report	Supervisor's Initial Class Report	MR - S1A	S - ICR
Monthly Report I	Supervisor's Monthly Report Part 1	MR - S2A	S - MR1
Monthly Report II	Supervisor's Monthly Report Part 2	MR - SB	S - MR2
III. *Partner Agencies*			
First Month Report	Partner Agencies Initial Class Report	MR - PAI	PA -ICR
Monthly Report I	Partner Agencies Monthly Report Part 1	MR - PA III	PA - MR1

NEW TITLES AND CODE NUMBERS FOR THE INSTRUMENTS OF THE 1989 VERSION

Earlier titles	New titles	Earlier Code	New Code
Monthly Report II	Partner Agencies Monthly Report Part 2	MR - PA II	PA - MR2

[Copies of all reports to District Coordinators]

IV. *District Co-ordinator's Forms*

Earlier titles	New titles	Earlier Code	New Code
Monthly Report	District Monthly Report Part 1	MR - POA	D - MR1
Project Area Progress	District Monthly Report Part 2	MR - POB	D - MR2
Training Activities	District Periodic Report on Training Activities and Training Needs	MR - POC	D - TR

[DCs fill in a special D-MR 1 and D-MR 2 for Partner Agencies and send them to RCDOs]

V. *Regional CDO Forms*

Earlier titles	New titles	Earlier Code	New Code
-	Regional Monthly Report Part 1	-	R - MR1
Monthly Report for Monitoring etc.	Regional Monthly Report Part 2	MR - ROA	R - MR2
-	Regional Monthly Report on Training Activities and Training Needs	-	R - MTR

[Regional Officers fill in a special R-MR 1 and R-MR 2 for Partner Agencies and send them to Headquarters]

NEW TITLES AND CODE NUMBERS FOR THE INSTRUMENTS OF THE 1989 VERSION

Earlier titles	New titles	Earlier Code	New Code
VI. *National Level Headquarters*			
Annual National Progress by Month	National Annual Progress Report by Districts	Table 1	N - T1
Monthly National Progress by Project	National Quarterly Report cum Annual Progress Report by Regions	Table 2	N - T2
Training Activities (Annual Summary)	National Report on Training Activities and Training Needs	Table 3	N - TR

[R-MR1 will not be aggregated at National level but will be collected and acted on by Headquarters]

VII. *Additional Forms*

Guidelines for Supervision (to be used as checklist by Supervisor visiting a literacy class).	S-GS
Radio Group Reports (to be filled in by Instructor and sent via Supervisor to DC, to RCD, to Radio Section at Headquarters).	I - Radio

Neither Literacy Tests to be administered to learners, nor the scoring sheets developed for marking tests have been included in the above list. Test data would, of course, appear in various forms and reports that make up the MIS.

(The above list is reproduced from National Adult Literacy Programme, *Guidelines and Instruments for Monitoring the Programme* (Third Revised Edition). Lilongwe: National Center for Literacy and Adult Education, Ministry of Community Services, Government of Malawi, 1989.)

The flow of information across levels is indicated in Figure 3 on the next page.

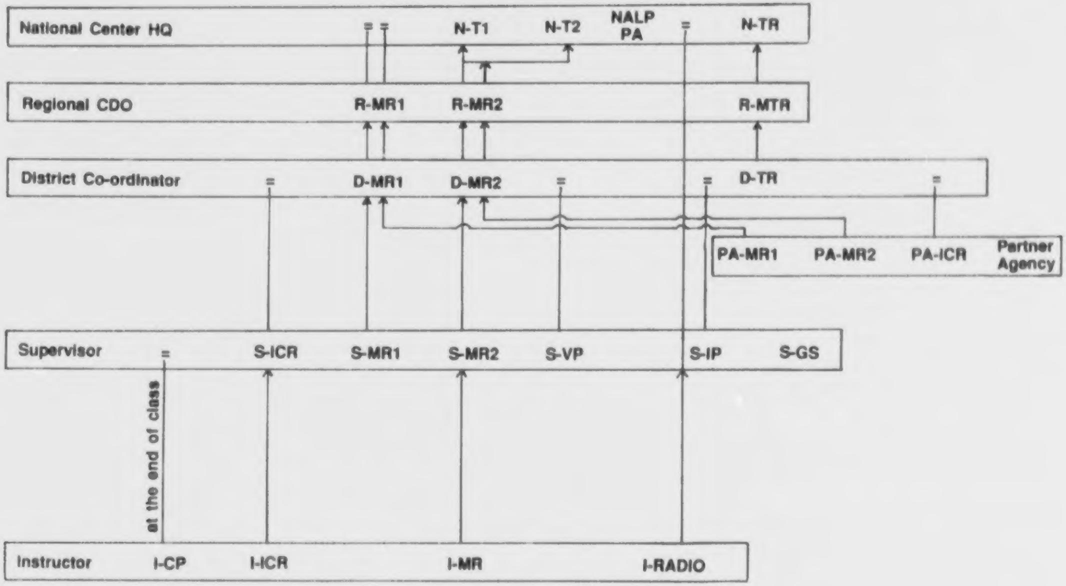

National Center HQ
N-T1
N-T2
NALP PA
N-TR
Regional CDO
R-MR1
R-MR2
R-MTR
District Co-ordinator
D-MR1
D-MR2
D-TR
PA-MR1
PA-MR2
PA-ICR
Partner Agency
Supervisor
S-ICR
S-MR1
S-MR2
S-VP
S-IP
S-GS
at the end of class
Instructor
I-CP
I-ICR
I-MR
I-RADIO

Figure 3: MIS Information Flowchart

vii. *Labelling, coding and numbering of instruments*

The participants of the proposal writing workshop had identified several problems with labelling and coding of forms, and the proposal for revision of the MIS sought to remove those problems. Chief among them was the lack of uniformity in labelling and coding. A careful examination of the list of forms above should clarify how the new numbering system solved the previous problems.

viii. *Printing and distribution*

The final hand-made forms of various instruments were made as close to the final product as possible and "dry runs" were conducted to see if those forms could be easily filled. The proposal for the MIS did not, however, include any suggestions about printing, a matter which was once again left to administrative decisions. The questions of distribution among various regions and districts and classes were discussed, but did not become part of the proposal.

ix. *Training of functionaries*

All participants were pleased that they had had the opportunity of coming to the workshop. While they were satisfied with the opportunity to contribute to the revision of the MIS, they were even more gratified with the "training" they had received in the maintenance and utilization of the MIS.

They were all convinced that the proposal should include recommendations for the training of all functionaries -- instructors, supervisors, and staff at district, regional and national offices in the maintenance and use of the MIS. Indeed, requests were to be made to a donor agency for assistance in the conduct of training workshops to cover each and every functionary in the system.

x. *Location of records*

The question of location of records was again given less attention than it deserved. The general sense was that when records were no longer needed in the localities, districts or regions, they should be sent to the center. Not only would those records be safer there, they could also be used to develop special studies on the basis of random selections of participants from the national pool.

Things to do or think about

1. Review quickly but carefully all the files and records that are already being kept in your office on the adult literacy project, program or campaign. What steps would be necessary to change the existing "files and records" into a more systematic MIS?

2. Think of a minimum set of tables of literacy statistics that you would like to include in your quarterly reports each time you write such a report as the director of the literacy project.

3. Is the MIS for Malawi described above adaptable to your setting? What can be borrowed, and what cannot be?

CHAPTER 7

THE PROCESS AT A GLANCE: TOOLS AND TECHNIQUES OF IMPLEMENTING AN MIS

To implement and install an MIS, skills are needed in such techniques as concept analysis, indicator writing and development of levels and standards; construction of survey instruments and checklists, achievement tests and attitudinal scales; and development of observation schedules. Some ideas about developing filing systems for records on paper, and on film and tape should also be learned.

In implementing and installing an MIS for a literacy campaign, program or project, we shall need to do more than count the numbers of participants, and record their age and gender. We have to deal with much more complex concepts, such as a "dropout", "reading ability", "teacher effectiveness", "adoption of innovation", "development" and so on.

These are concepts that cannot go directly into the rows and columns of tables designed for our MIS. Before we get to asking, seeing, counting and recording, these concepts will have to be "unpacked" through a process of *continuous elaboration* involving more than one cycle of *concept analysis.*

The process of elaboration does not stop with concept analysis. After analysis, concepts may need to be defined in terms of *indicators* -- observable happenings and behaviors which would indicate the likely presence of something not visible to the naked eye.

For each of the indicators, one or more *items* (questions to be answered, choices to be tick-marked, blanks to be filled) may have to be written to learn enough about the same one indicator. These items will then have to be organized into *instruments* that can be taken to the field and used in the process of data collection. These instruments may take the form of surveys, interviews, questionnaires, tests of knowledge and performance, attitude scales, checklists, and observation schedules.

After collecting data, the instruments may again have to be broken down into items for inclusion in the MIS. To answer different questions raised by policy-makers, planners, managers and program people, items of data in the MIS will have to be combined in different ways to come up with needed answers. The process may require weighting of scores, establishing of intervals, levels and standards, and statistical treatment of data.

Figure 4 on the next page presents the total process at a glance. Detailed discussion of some of the steps listed above follows.

SECTION A: Concept Analysis

Evaluation needs arise and evaluation tasks are *initially* stated in rather general terms. For example, the evaluation objective may be to "evaluate the impact of a literacy or a post-literacy program on a community". "Impact", of course, is not something we can evaluate in one gulp! We have to do a "concept analysis" of the concept -- impact -- in the particular context of a literacy or a post-literacy program in a particular country. Concept analysis is *analysis*, which in its dictionary meanings is the "separation or breaking up of a whole into its fundamental elements or component parts; a detailed examination of anything complex made in order to understand its nature or to determine its essential features".

The questions we are asking in each concept analysis are: What are the generic meanings of the concept we are analyzing? What are *not* the meanings of a concept? How can the concept be differentiated from other similar concepts? For example, will "impact" be differentiated from "unanticipated consequences" of literacy or not? A conditions-type analysis can also help: What conditions must be met for a literacy worker to claim something to be the impact resulting from a literacy program? Finally, the question must be asked: What should be happening in terms of "operations" for a particular concept to be manifest in the real world? This is what is often referred to as *operationalization*. In other words, the concept is defined in terms of operations, happenings, actions, behaviors and things that have concrete existence.

Concept analysis need not be a purely logical process conducted by an expert. We consider sociological strategies for concept analysis to be as important as logical strategies. Analysis of concepts like literacy impact must, at some stage, go through a participative process of definition. In this way, it will be possible for the various stakeholders to project their plural values in regard to what they want literacy or post-literacy to do or to have done for them.

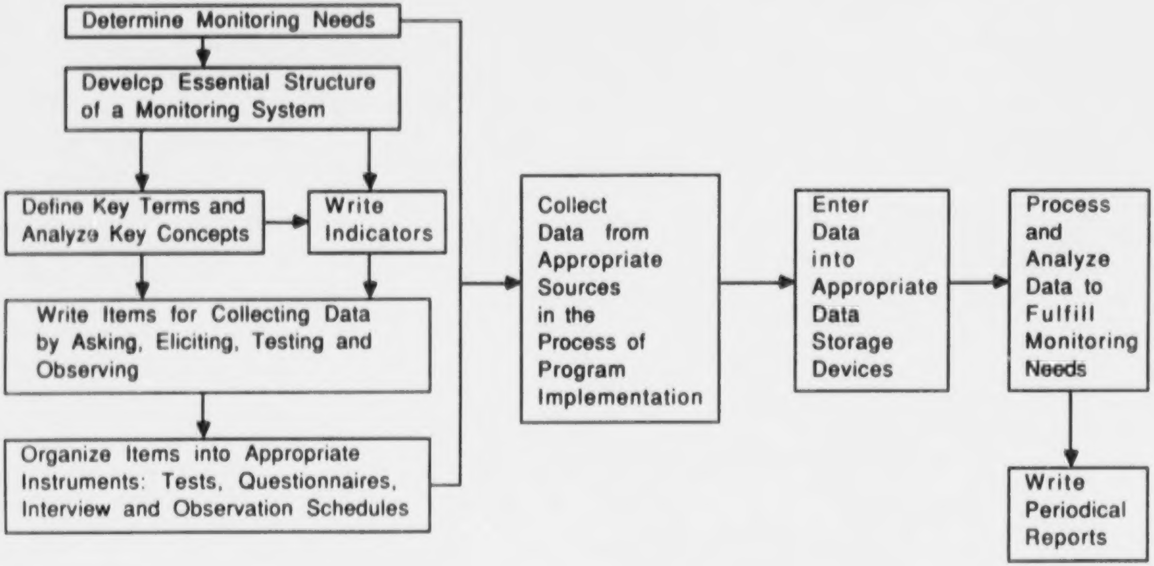

Determine Monitoring Needs
Develop Essential Structure of a Monitoring System
Define Key Terms and Analyze Key Concepts
Write Indicators
Write Items for Collecting Data by Asking, Eliciting, Testing and Observing
Organize Items into Appropriate Instruments: Tests, Questionnaires, Interview and Observation Schedules
Collect Data from Appropriate Sources in the Process of Program Implementation
Enter Data into Appropriate Data Storage Devices
Process and Analyze Data to Fulfill Monitoring Needs
Write Periodical Reports

Figure 4: The Process of Development and Utilization of a Management Information System (MIS)

Analyzing the concept of "literacy impact"

Using some of the purely logical strategies of concept analysis (to be later complemented by sociological strategies in participation with all stakeholders in the field), we can conceptualize literacy impact as having the fundamental elements and essential components implied in the following grid.

GENERAL DIMENSIONS OF IMPACT

	Socio-cultural	Political	Economic	Environmental
Loci of impact:				
Individuals	1	2	3	4
Groups	5	6	7	8
Institutions	9	10	11	12
Community/ Sub-culture	13	14	15	16

There are, of course, other ways of categorizing the dimensions of the impact of literacy and post-literacy on the lives of people. Literacy workers today like to talk of three dimensions of impact, relating to (1) literacy skills, (2) functionality, and (3) awareness--a term which is now being used in preference to conscientization or consciousness-raising.

It is possible, of course, to use the above grid to develop a list of elements and components *separately* for each of the multiple dimensions of literacy impact. In real life, literacy workers will choose practical rather than theoretical analyses. It should be interesting to see a list of components and elements that appeared in a concept analysis of the impact of literacy and post-literacy in a document of the Unesco Institute for Education. This is seen in Example 1 on the next page.

We fully accept and strongly endorse the concept of *generalized* functional literacy that includes literacy skills, functionality and awareness. But we realize also that literacy skills are central to a literacy program. This being so, we focus on literacy and post-literacy to exemplify the continuous nature of concept analysis.

EXAMPLE 1

On the basis of case studies of ten countries, the literacy, post-literacy and continuing education programs were known to have produced impact of the following nature:

(a) changes in attitude
(b) changes in occupational skill and income
(c) changes in personal habits, hygiene, family planning, etc.
(d) future orientation, aspirations
(e) increase in school enrollment
(g) community support to construct schools
(h) communities undertaking responsibility for their own goals
(i) emergence of new community leaders
(j) trained personnel for adult and adolescent education
(k) increased use of documents for civic purposes
(l) increased use of state services
(m) better nutritional practices
(n) improved village sanitation
(o) generation of achievement-oriented attitude
(p) demand for better communication system
(q) increased self-confidence and awareness
(r) involvement in the social environment
(s) command and transformation of environment
(t) levels of knowledge attained by learners in the areas of vocational training, health and civics
(u) participation in marketing, supplies and agricultural extension in the village
(v) application of literacy skills for individual and community purposes
(w) changes in the technology used by people
(x) changes in the structures servicing people
(y) greater political awareness and participation in decision-making
(z) more knowledge of modern ways of farming and creation of favorable literacy environment

---

[Draft Report of Groups A and B (PRG 4.32 Review Meeting, 1984). Unesco Institute for Education (UIE) project on the Development of Techniques and Procedures on Evaluation pertaining to Programs of Literacy and Post-literacy in the Framework of Lifelong Education, Hamburg, June 24-28, 1985.]

Since impact of literacy programs must appear in the form of learning of literacy skills by adult learners, what elements and components of literacy skills shall we be looking for to be able to say that literacy or post- literacy has manifested itself in the lives of people and communities? Once again, examples are offered from actual literacy and post-literacy programs from Third World countries.

We begin with a somewhat simple description, in Example 2 on page 112, of literacy skills. The project organizers label it "Evaluation Criteria".

We offer it as an example of "concept analysis" of the concept of literacy. We should note that these concept-components or criteria can easily be used to construct a literacy test. The score on such a test will become an "indicator" of success or failure of a learner in his/her performance in literacy.

The concept analyses of literacy and post-literacy developed within the Indian National Adult Education Program, and shown in Examples 3 and 4 on pages 113-114, were much more elaborate than those presented in the example of the literacy campaign in Kerala.

EXAMPLE 2
THE TLM LITERACY EVALUATION CRITERIA

I. *Reading*:

A. To read out aloud any passage of the learner's choice, correctly pronouncing 30 words a minute.

B. To read to oneself any simple worded book (that the learner is not familiar with) at the speed of 35 words a minute.

C. To read and understand road signs, posters and other simple instructions.

D. To read simple literature in his/her own work environment.

II. *Writing*:

A. To understand the meaning and copy out a passage at 7 words a minute.

B. To listen to dictation and write at a speed of 5 words a minute.

C. To write legibly and leave correct spaces.

D. To be able to write simple letters and messages and also fill up simple forms required in daily work.

III. *Arithmetic*:

A. To write and read from 1 to 100.

B. To do addition and subtraction of up to 3 digits and multiplication and division of 2 digits.

C. To understand metric system of weights and measures, money, distance, area, and to tell the time.

D. To have rudimentary knowledge of proportion and interest.

[Source: Center for Development Studies, *Report of the Total Literacy Mission*, CDS, Trivandrum, Kerala, India, c. 1990, p. 43.]

EXAMPLE 3
SPECIFICATION OF NORMS FOR LITERACY ATTAINMENT

- a) *Reading Skills*
  - i. The learner should, at the end of the programme, be able to read correctly a simple passage of about five to six sentences in a minute. Such a passage may be from the reading material used at the centre and should be preferably in the same letter type.
  - ii. The learner should be able to read approximately 10-20 words of hand-written (bold) material, per minute.
  - iii. The learner should be able to read with understanding road signs, posters, simple instructions, and some headlines of newspapers for neo-literates.
  - iv. The learner should be able to read figures from 1 to 100.
  - v. The learner should be able to comprehend the material read in items i, ii, iii above and should be able to answer questions relating to it.
- b) *Writing Skills*
  - i. The learner should be able to copy out a minimum of ten words per minute from a small passage. The words in the passage may be of not more than four letters. He/she should also be able to understand what is written.
  - ii. The learner should be able to take down dictation at the speed of at least seven words per minute.
  - iii. The learner should be able to write in a straight line with proper spacing on ruled paper.
- c) *Computational Skills*
  - i. The learner should be able to make minor calculation of up to three digit figures involving simple addition, subtraction, multiplication and division. The divisor in the case of division and the multiplier in the case of multiplication should be one digit.
  - ii. At the end of the course the learner should be in a position to gain a practical knowledge of metric weights and measures.
  - iii. The learner should know tables up to 10.
- d) *Application of Literacy Skills*
  - i. The learner should be able to read captions, signboards (written road-signs), posters, newspaper headlines, and other communications that come to him in legible and bold handwritten papers.
  - ii. The learner should be able to write simple letters, simple applications, and fill up forms such as money order, loan and bank forms.
  - iii. The learner should be able to keep accounts of day-to-day expenditure and savings and be able to check entries in his/her post office or bank pass-book.
  - iv. The learner should be able to follow and act upon instructions given on bags of fertilizers, pesticides, seedicides and medicines, etc.

EXAMPLE 4
LITERACY AND NUMERACY COMPETENCIES TO BE ACHIEVED ON COMPLETION OF POST-LITERACY STAGE

1. *Language*

i. Speaking
- Ability to participate in discussion
- Ability to describe experiences

ii. Reading
- Ability to read aloud and fluently, simple printed material with correct pronunciation, intonation and stress
- Ability to read silently and with a speed of 70 words per minute
- Ability to read a variety of printed material with comprehension (stories, informative material, text-books, notices, newspapers, posters, various forms, etc.)
- Ability to read hand-written material (letters, messages, instructions, etc.)

iii. Writing
- Ability to copy with understanding at a speed of 15 words per minute
- Ability to take dictation at the rate of 10 words per minute
- Ability to write with understanding simple messages
- Ability to write independently letters, applications, and to fill up forms for bank loans, money-order, etc.

2. *Numeracy*

- Ability to read and write numbers (up to 10,000)
- Ability to compare and arrange numbers (up to 10,000)
- Ability to understand the concept of place value of numbers (up to 5 digits)
- Ability to solve sums involving addition of two or more numbers (the total sum not exceeding 10,000)
- Ability to solve sums involving subtraction of one number from another (up to 4 digits)
- Ability to solve sums involving multiplication of a number by another number (the multiplier being up to 2 digits)
- Ability to solve sums involving division of a number by another number (the dividend being up to 4 digits and divisor up to 2 digits)
- Ability to solve problems involving 2-3 operations (using not more than 4 digits at any stage of the operation
- Ability to use unitary methods, calculate simple interest, percentage, etc.
- Ability to do simple calculations involving standard units of currency, time, measurement, weight, area, volume, etc.
- Ability to maintain accounts and solve day-to-day problems involving numeracy

[Examples 3 and 4 are from DAE, *Report of National Workshop on Monitoring and Evaluation.* New Delhi: Directorate of Adult Education, Ministry of Education, Government of India, 1982. Also, reproduced in Mathur, R.S., "Evaluation of literacy and post-literacy programs in India", (PRG 4.32/4.38), Unesco Institute for Education (UIE) project on the Development of Techniques and Procedures on Evaluation Pertaining to Programs of Literacy and Post-Literacy in the Framework of Lifelong Education, Hamburg, October 7-11, 1985.]

Levels and intervals

As we have stated earlier, a continuous process of elaboration and further elaboration is involved, as we go from general concepts through concept analysis and indicator writing to item writing for evaluation instruments. A careful examination of Examples 1 to 4 should show that some components of the concepts analyzed above would require further analysis before indicators can be written and items can be constructed. Some other components may be ready for indicator writing and construction of items.

It should be noticed from Example 3 and Example 4 above that a hierarchy of skills has been built into the lists of criteria. Reading, writing and numeracy skills required at the post-literacy stage are, of course, higher than those required at the initial literacy level. Even within each of the two lists of components there is an implicit hierarchy. In other words, intervals of achievement or levels of performance have been built into the lists themselves. Examples 5 and 6 on pages 116-118 identify literacy levels more explicitly.

EXAMPLE 5
LEVELS OF ACHIEVEMENT IN FUNCTIONAL LITERACY IN TANZANIA

To avoid categorizing learners as pass or fail, it is possible to grade them according to level of difficulty of response required.

Level I:
A learner who has enrolled and has attended two thirds of the literacy sessions in any one year of literacy activities.
Level II:
A participant who qualifies for Level I above, but who has also successfully passed one or both tests in the following sub-levels:
Sub-level (i):
A person who is able to recognize words and/or symbols, writes letters of the syllables, writes numbers and/or arithmetic signs including simple mental calculations.
Sub-level (ii):
A person who is able to read a short, simple meaningful sentence, who is able to write a simple short sentence and can add and subtract one figure numbers.
Level III:
A person who qualifies for Level II above, but who has also successfully passed one or both tests in the following sub-levels:
Sub-level (i):
A learner who is able to read a short, simple meaningful sentence, who is able to write a simple short sentence and can add and subtract two-figure numbers.
Sub-level (ii):
A person who possesses mastery over symbols in their written form, or is able to encode and decode written messages. Such a person should be able to perform the following: to read fluently a simple text with understanding (the text itself being based on common syllables and vocabularies in the functional primers and according to the most frequent syllables and vocabularies used in the Swahili language). He should also be able to write a simple short message or passage; add and subtract three-figure numbers, multiply two-figure numbers, and divide by one figure.
Level IV:
A learner who continuously uses the acquired literacy skills. Such a person should have qualified in Level III above, but also should be able to read and write messages; be able to read a newspaper (for example, *Uhuru*, *Ukulima wa kisasa*, etc.) to keep up with current happenings and obtain information; be able to read "how to do it yourself" books, little books on better living, better food, better ways of farming, etc.; and be able to keep records and solve simple arithmetic problems. He should also be able to keep a simple book of accounts on income and expenditure.

Those participants who achieve Levels III and IV in reading, writing and arithmetic combined are considered literate graduates and those participants who achieve Level IV are considered functionally literate.

EXAMPLE 6A

Table 1: RESULTS OF THE SELF-EVALUATION TESTS

A. Results defined and detailed by level

Subject	Level	Required Skills
READING AND WRITING	0	Practically no literacy skill
	1a	Read aloud correctly words of one or two syllables chosen from those taught in the first three lessons of the textbook
	1b	Write words of this difficulty from dictation
	2a	Read aloud correctly sentences of five to nine words based on the syllables taught in the first eight lessons of the textbook (two thirds of the phonemes of the written language)
	2b	Write sentences of this difficulty from dictation
	3a	Read aloud correctly a short passage of three or four sentences, each sentence comprising up to 12 words composed of any syllables of the written language
	3b	Write a short passage of this difficulty from dictation
	4a	Read a text of several paragraphs and reply correctly to questions relating to it
	4b	Write an essay of several paragraphs

EXAMPLE 6B

Table 1: RESULTS OF THE SELF-EVALUATION TESTS

B. Results defined and detailed by level

Subject	Level	Required Skills
ARITHMETIC	0	Practically no knowledge
	1a	Read and write the digits 0 to 9
	1b	Read and write three-figure numbers
	2a	Do addition, with carrying, of two or three numbers of three, four or five digits
	2b	Do subtraction, with carrying, of numbers of two or three digits
	3a	Multiply, with carrying, numbers of two or three digits
	3b	Divide numbers by two or three digits with remainder
	4a	Solve dictated practical problems involving division by two- or three-figure numbers with remainder
	4b	Solve dictated practical problems involving a series of operations including a division of level 4a

---

[Examples 6A and 6B are both taken from Ouane, A., "Evaluation and Monitoring of Literacy and Post-Literacy Programs in Mali--The Experience of DNAFLA (National Directorate of Functional Literacy and Applied Linguistics)", PRG 4.38, Doc 5, Unesco Institute for Education (UIE) project on the Development Techniques and Procedures of Evaluation Pertaining to Programs of Literacy and Post-Literacy in the Framework of Lifelong Education, Hamburg: Unesco Institute for Education, March 1989.]

Examples 1 - 6 do not by any means exhaust all the concept analysis tasks of literacy workers. Seemingly simple concepts such as "dropouts" will have to be defined in the context of particular programs. Questions such as these will have to be answered: How do we make sure that dropouts do not include no-shows (those who registered for classes but did not actually enroll); pushouts (learners pushed out of groups by teachers because they were left too far behind in the group); or successful completers (those who feel they have achieved what they had set out to achieve and consider it unnecessary to continue)? How do we treat a learner who dropped out for a period of time and then came back?

Before taking leave of the topic of concept analysis, we must deal with a few other important concerns of literacy workers -- teacher effectiveness, and curriculum evaluation.

What is effectiveness? It will, of course, have to be defined in the context of each particular program under evaluation. Example 7 on page 120 will demonstrate the complexities involved in the concept of teacher effectiveness.

In Examples 8 and 9 on pages 121-122, we turn to the issues of curricular evaluation.

EXAMPLE 7
EFFECTIVE TEACHER

A. *Knowledge of subject matter*
Is the teacher knowledgeable?

B. *Organization and clarity of presentation in the group*
Is the teacher's presentation structured? Is there self-conscious use of teaching strategies? Are important points summarized?

C. *Instructor-learner interaction*
Is there discussion in class? Is everyone encouraged to ask questions? Are the teacher's questions stimulating?

D. *Level of enthusiasm*
Does the teacher show enthusiasm for teaching? Does the teacher show respect for learners?

E. *Use of instructional materials*
Are instructional materials in use? Are some of these produced locally? Is there use of indigenous media and institutions in instruction?

F. *Use of extension workers in teaching*
Is there collaboration with other extension workers in the field?

G. *Achievement of learners*
Are adults learning? What? How effectively?

H. *Provision of feedback to learners about their progress*
Does the teacher let learners know how they are doing, honestly, but tactfully?

I. *Help in transfer of learning to life outside the class*
Does the teacher help in the transfer of learning to life outside the class? Does the class have an income-generating project?

EXAMPLE 8
EVALUATING CURRICULUM EFFECTIVENESS

The concept of curriculum effectiveness can be concept analyzed to have the following components:

1. General appropriateness
2. Built-in possibilities for the identification of errors
3. Feasibility
4. Quality
5. Standards
6. Utility
7. Adequacy
8. Relevance
9. Responsiveness to learners' needs
10. Appropriateness of content and method
11. Internal consistency
12. Clarity
13. Suitability to program objectives
14. Up-to-date-ness
15. Balance
16. Avoidance of breakdowns in teaching by anticipating difficulties faced by learners

---

[Unesco Institute for Education (UIE) materials. First draft of cross-national synthesis (PRG 4.32): Learner evaluation, curriculum evaluation, program monitoring and impact evaluation, Hamburg, June, 1985.]

EXAMPLE 9
EVALUATING FOLLOW-UP BOOKS THROUGH PEER REVIEW: DIMENSIONS/COMPONENTS OF GOODNESS

1. Size of the book
2. Quality of paper
3. Binding and general presentation
4. Typefaces and type sizes
5. Leading and spacing
6. Title page -- exactness, attractiveness, display
7. The intent of the book
8. Construction of argument
9. Unity and coherence of content
10. Quality of illustrations and integration of the verbal and the graphic
11. Quality of the message
12. Literary treatment and writing style
13. Readability level
14. Chapter and paragraph division
15. Spelling, punctuation and typographical errors
16. Overall impression

[From Bhola, H.S., *Writing for New Readers: A Book on Follow-up Books* (Revised Version), Bonn: German Foundation for International Development, 1984.]

SECTION B: Writing Indicators

As we have mentioned repeatedly, the general process of elaboration from general concepts to test items can involve several cycles and repetitions of concept analysis and indicator writing. After some abstract concepts such as literacy, post-literacy, teacher effectiveness, efficiency and motivation have been analyzed (unpacked, and their different parts specified), another problem arises: How do we know that these abstract things actually *exist* in the field and are *changing* by some degree, in some direction? We should remember that it is not always possible to state clearly where concept analysis ends and indicator writing begins. While the two are useful analytical categories, they can be quite ambiguous in the real world of work.

The problem is that many concepts and their components such as individual motivation and commitment, problem-solving capacity, political awareness, community cohesiveness, responsiveness of social institutions, and the quality of life are not visible to the "naked eye". We will need some concrete manifestations of behavior, some signs, which will indicate the presence of high motivation, problem-solving capacity, community cohesiveness and responsiveness of institutions. These signs are what we call indicators.

The process of developing indicators is complex, to say the least. Indicators must be valid, they must be concrete, and they must be parsimonious (that is, the list of indicators for a condition should not be impractically long). To be able to engage in indicators research, one must have a good enough understanding of the behavior of individuals, institutions, and societies. It would be ideal to have sufficient grounding in logic and social science theory. We cannot, however, wait for ever to become expert social scientists. As practitioners and evaluators, we must learn to develop good enough indicators. As in the case of conceptual analysis discussed in the earlier section, participative strategies can also be used in developing indicators.

An introduction to indicators research

Indicators research has emerged as an important area of research in its own right over the last twenty years. A brief introduction to some of the traditions of indicator research will be useful at this stage:

Economic indicators

Economic indicators have been the oldest and most frequently used. Most of us are familiar with the Gross National Product (GNP) per capita, the most widely used economic indicator. Interest rates and rates of inflation are other economic indicators.

Social indicators

In recent years, considerable attention has been given by social scientists to social indicators. The social indicators of the wellbeing of a family, for instance, have included cash income, net worth of assets owned by a family, a family's endowment of human capital, the variability of income over time, intrafamily transfers of income, the impact of government expenditures and taxes, and leisure and nonmarket productive activities. All of these are more concrete components and indicators of the more abstract concept of the family's wellbeing.

Health indicators as social indicators

Health indicators can be seen as a special class of social indicators. Life expectancy, infant mortality, population per physician, percentage of population with access to safe water, daily per capita calorie supply as percentage of requirement, are some of the typical health indicators. While most of these indicators relate to countries, they are transferable for use at the regional and community levels as well.

Science indicators as social indicators

Science indicators are also social indicators since they indicate the level of science and technology in a society. The number of science students in the science track in secondary schools, the number of scientists produced by universities, the number of patents awarded for original scientific inventions, the trade balance in technology-intensive products, and the research and development (R&D) expenditure by the government as percentage of gross GNP, are examples of science indicators.

Educational indicators

Indicators research in the area of education is now attracting more and more attention. The list of indicators developed by Gooler[1] is shown in Example 10 and should be of interest to readers.

EXAMPLE 10
CATEGORIES OF EDUCATIONAL INDICATOR

Access
How many and what kinds of people participate in educational activities
Retention rates in educational activities
Catalog of existing/available educational activities or services

Aspirations
Description of needs and desires of various kinds of people
Individual self-assessments of personal capabilities
Description of institutional goals

Achievement
What people know, do, and feel
What people have earned (degrees, diplomas, certificates)
What is taught

Impact
Consequences of having schooling
Impact of education on social/economic/cultural systems
Consequences of not having schooling

Resources
Capital, personnel, and material expenditures
Quality of human resources
Cost to benefit/effectiveness ratios
Quality of educational climate
Time

Note that most of the listed indicators are numbers or are at least nominal categories -- High, Medium, Low, etc.

Indicators of indicators

As can be noticed, some of the "indicators" listed above as economic, scientific, health, social and educational indicators are in fact quite general concepts, too broad to be usable for data collection. (Remember: One person's concepts are another person's indicators, and vice versa!) In fact, we often have to go through a multi-step process of developing indicators of indicators, and indicators of "indicators of indicators".

In the above, we have tried to define indicators by exemplification. We have tried to show what indicators in various areas of economics, science and technology, health and education look like. Most of this indicator research has been done at the national and international levels. The evaluator can sometimes make direct use of the indicators developed at these levels. More often, however, the evaluator will have to develop indicators that make sense in his or her concrete situation.

Indicators of clear and direct interest to literacy workers

The indicators of interest to an evaluator will relate to literacy skills, functionality and awareness at the individual level, and to developmental impact on groups, institutions and communities at the macro level. It is not possible within the scope of this handbook to actually work out the indicators for all of the needs of literacy workers. It should be kept in mind that standardized sets of indicators are not possible because indicators writing has often to be done in terms of a particular content; and relative to the concept definitions developed in a particular program setting. The general principle is to go from a first abstraction through categories and subcategories to behavioral manifestations: to things which can be seen, heard, touched, sensed, judged and scored.

Validating indicators

Unfortunately no standard formulas can be suggested for writing indicators and testing their goodness - their reliability and their validity. Ultimately the goodness of indicators will be proven through their testing-by-use. It would always be a good idea, however, for literacy workers, trainers and development agents to pre-test their indicators through peer reviews. They should show their indicators and their "indicators of indicators' to their colleagues and let them criticize their work.

The indicators-instrument connection

We shall be discussing evaluation instruments and their construction and use later in the chapter as well as in other chapters to follow. However, we wish to return here to Figure 4 at the beginning of this chapter showing the process of MIS design at a glance. The clear and direct connection between the process of indicator development and the process of instrumentation should be noted. In the construction of tools and instruments, we merely take the next logical step from indicator development. We ask: What data or evidence should be collected to demonstrate the existence of or change in the indicator-related behavior or condition? How do we elicit and collect the required data or evidence? What aids (tests, tapes, questionnaires, schedules, etc.) might be used for recording the data or evidence?

How are some of these instruments -- achievement tests, attitudinal scales, interview and observation schedules made? We now turn to these concerns.

SECTION C: Making Tests of Achievement

It is important here, once again, to return to Figure 4 on page 108 showing the process of MIS design at a glance. Once the evaluation concerns and questions have been stated, key concepts identified and analyzed, and indicators worked out, it is time "to go into the world" and to look for evidence. There are a limited number of things one can do to make reality unfold, to make the world give away its secrets. These are the choices we seem to have:

We can *see*, or *observe*
    As participants or non-participants
        -- overtly or covertly
We can *ask*
    The person concerned
        -- directly or indirectly
    Someone other than the person concerned
        -- directly or indirectly

We can *elicit* behaviors and then *record* those behaviors
-- overtly or covertly
We can *read*
-- documents or tell-tale signs
We can *count*

The evidence can be recorded with the help of a variety of items. These items can then be organized into instruments which may be structured (tests, questionnaires) or unstructured (journal or diary).

Scales of measurement

Items we write for our instruments are in reality items of measurement. Thus, measurement is the essence of most information gathering, especially in an MIS and RE. We often need to go beyond crude comparisons in terms of good, better, or best; or big, bigger or biggest. To do this we need standard yardsticks with which we can take the measures we want; and can state *how much* of a difference exists between two entities, and in what direction.

Unfortunately, in the social sciences we do not have the benefit of such tools as micrometers, carbon dating and atomic clocks. Our measures and yardsticks are often quite crude. We need to understand, however, the nature of scales that are available to us; and we need to understand their possibilities and their limitations.

The nominal scale

The nominal scale does not really measure, it only nominates objects to categories. The classification of adults in a community into males and females, and assigning them numbers (1 for males, 2 for females), will be an example of using a nominal scale.

We need to understand that numbers used in the nominal scale mean nothing in regard to the value of categories, except to show that they are different. In the above example, 2 (for females) is not twice 1 (for males). The numbers 2 and 1, in this particular context, cannot be added or subtracted from each other in any meaningful way. They merely serve as codes.

The ordinal scale

The ordinal scale introduces ordering to the nominal scale. The categories can now be ranked in an order of succession as "First, Second and Third" or "Good, Average and Poor".

The ordinal scale, again, could be assigned numerical values: for example, 5 for Good, 3 for Average, and 1 for Poor. But, once again, 5 is not five times 1 in terms of the scale, nor is 6 (two Averages) better than 5 (one Good).

The interval scale

The interval scale, as the name suggests, has intervals which make mathematical sense. On a meter rod, the difference between 3 and 5 centimeters is the same as the distance between 53 and 55 centimeters.

Scores on an achievement test are in reality ordinal data, but we can often treat them as if they were interval data. We can say that B made twenty points (or twenty intervals) more than A. However, if B had made 40 points and A had made 20, we could not say that B is twice as good as A. To be able to make that kind of statement, we will need ratio scales.

The ratio scale

The ratio scale, in addition to being an interval scale, has an absolute zero. This means that 25 is 5 points more than 20, and that 60 is three times as good as 20. Thus, the ratio scale permits us to work out ratios and proportions. Two meters is twice as long as 1 meter. One thing can be twice as hot as another.

We need to keep the properties of various scales of measurement in view as we deal with data from our various evaluation tests, tools and instruments.

Organizing items into instruments

Clearly, it would be silly to jumble all the items of measurement together somehow, in some sequence, in the same instrument. Various considerations of completeness, logic, socio-logic and simple convenience are involved:

It is often better to separate observation items from interview questions; and to separate interview questions from test items. (We are not saying, however, that these combinations are not possible or should never be attempted).

It is desirable to bunch together similar items and to sequence items or clusters of items in such a way that there is a logic to the total instrument -- the logic of meaningful conversation in the interview schedule or questionnaire, and the logic of "simple to complex" in an achievement test.

It is better to make each separate instrument self-contained by including appropriate demographic items so that it can be interpreted properly in a particular context.

You may be able to think of some other considerations that must enter instrument design.

How to administer instruments

A whole body of experience has become available in literature on how to administer tests and other instruments to our respondents. It is impossible to treat this subject with any completeness within the scope of this monograph. Only a few key ideas can be presented:

- The need for a relationship of equality, mutuality and trust between the evaluator and the respondent
- The need for proper explanations of purposes and modes of response without leading the respondents to give particular types of answer
- The creation of conditions of convenient access, privacy, quietness, and personal comfort for the respondent to provide responses

Achievement tests

An MIS would often include data obtained from tests of knowledge, attitudes and performance. Sometimes, some interview and observation data will also be included in an MIS. In this section, we shall deal with tests of achievement as testing data most likely to be included in an MIS for a functional literacy program. It is important, therefore, that the evaluator be familiar with testing at the level both of theory and of methodological techniques.

Anyone who has been to school, has been subjected to tests (or exams as they are popularly called). Tests are a usual tool of the evaluator working with an MIS or what we have called the rationalistic paradigm. Tests, or achievement tests as they are often called, are tests of knowledge, skills and performance. Tests may be made to measure knowledge in arithmetic, biology, nutrition or animal husbandry; research skills, diagnostic skills or graphic skills; or actual performance in a role.

Tests can also be used to measure *aptitudes* (natural or acquired abilities or bents of mind). In fact, an aptitude test can be seen to be a special kind of achievement test.

Evaluators may sometimes be interested in testing *attitudes* (value dispositions and opinions). Attitude testing will be discussed later as part of attitude scales.

Having gone through many achievement tests in our lives and, perhaps, having ourselves written and administered tests as teachers and trainers, we might think of tests as relatively simple to make, to administer and to interpret. This is not really true. There are many complexities involved, as the discussion that follows would show.

Standardized norm-referenced tests and criterion-referenced tests

Tests may be made for one particular group (community health workers under training in a special workshop) or for a large regional or national population (all VIII grade students in Kenyan schools or even East African schools).

In the first case, the test will most likely be designed to measure whether the community health workers have achieved the criterion of success established in the particular context. The criterion of success may be a score of at least 80 out of the possible 100 on an achievement test specially designed for that group. This would be an example of a criterion-referenced test.

In the second of the two cases above, the test will most likely be designed to measure how well a student, a class, or a school is doing in comparison to other students, classes and schools tested on the same test of VIII grade mathematics or English or civics. To be able to make those comparisons, we will have to have norms -- how an average VIII grade student is supposed to perform on this particular test. When these norms do become available, the test becomes norm-referenced and standardized.

The process of standardization of tests for development of norms is itself quite standardized now. We do not discuss it here because trainers-evaluators will most often be dealing with criterion-referenced tests. Those are the tests we will focus upon in the following discussion.

Teaching objectives and testing objectives

Teaching and testing objectives should match with each other. It would be patently unfair to test learners on things they were never taught. This means that the test writer should have available to him or her a clear and detailed statement of the instructional objectives of a literacy course, to be able to make a test that will measure effectively the impact of the course.

Professors Benjamin S. Bloom,[2] David R. Krathwohl[3] and their associates have developed taxonomies of instructional objectives that should interest both literacy workers and test makers. The basic outlines of their taxonomies are reproduced on pages 133-134.

The test writer should not confuse the cognitive with the affective, or the ability to synthesize with the simple knowledge of universals and abstractions. We should realize that learning of information does not ensure real comprehension; and comprehension does not automatically lead to the ability to apply, analyze and judge. Similarly, it is possible to be positive *verbally* about a particular entity or a position without genuine commitment; and to have a set of discrete values that do not add up to a systematic and organized value system.

Choosing the test content

It is obvious that one cannot test everything that has been taught. One will have to take a small sample of all the knowledge taught, to be included in a test.

The sample of knowledge to be included in a test should be developed systematically from a detailed and comprehensive description of the subject matter taught. The two taxonomies presented above should be used for the description of subject matter taught: What factual knowledge was taught? What general principles and generalizations were communicated? What diagnostic skills and abilities to apply and transfer to other situations were underlined? What higher level processes were expected to be learned? What change in attitudes and values was reinforced?

Based on this comprehensive description, a sample of knowledge and values should be selected for test making.

INSTRUCTIONAL OBJECTIVES IN THE COGNITIVE DOMAIN

1.00 *Knowledge*
  1.10 Knowledge of specifics
  1.11 Knowledge of terminology
  1.12 Knowledge of specific facts
  1.20 Knowledge of ways and means of dealing with specifics
  1.21 Knowledge of conventions
  1.22 Knowledge of sequences
  1.23 Knowledge of classifications and categories
  1.24 Knowledge of criteria
  1.25 Knowledge of methodology
  1.30 Knowledge of the universals and abstractions in a field
  1.31 Knowledge of principles and generalizations
  1.32 Knowledge of theories and structures

2.00 *Comprehension*
  2.10 Translation
  2.20 Interpretation
  2.30 Extrapolation

3.00 *Application*

4.00 *Analysis*
  4.10 Analysis of elements
  4.20 Analysis of relationships
  4.30 Analysis of organizational principles

5.00 *Synthesis*
  5.10 Production of a unique communication
  5.20 Production of a plan, or proposed set of operations
  5.30 Derivation of a set of abstract relations

6.00 *Evaluation*
  6.10 Judgements in terms of internal evidence
  6.20 Judgements in terms of external criteria

INSTRUCTIONAL OBJECTIVES IN THE AFFECTIVE DOMAIN

1.00 *Receiving (attending)*
  1.1 Awareness
  1.2 Willingness to receive
  1.3 Controlled or selected attention

2.00 *Responding*
  2.1 Acquiescence in responding
  2.2 Willingness to respond
  2.3 Satisfaction in response

3.00 *Valuing*
  3.1 Acceptance of a value
  3.2 Preference for a value
  3.3 Commitment

4.00 *Organization*
  4.1 Conceptualizing a value
  4.2 Organizing a value system

5.00 *Characterization by a value or value complex*
  5.1 Generalized set
  5.2 Characterization

Types of test item
A variety of test items can be written to be included in an achievement test.

True/False. A statement is written and the respondent is asked to check it as true or false.

Example

Groundnuts and vegetables are
body-building foods T/F
(Answer: True)

True/False items are comparatively easy to write. These are, however, of limited use in testing for depth of understanding. The advantage of easy scoring is balanced by a disadvantage. Respondents feel encouraged to guess answers when they do not really know the answer. As they make guesses, they have a 50:50 chance of being right.

Short answer and completion items. As the name suggests, these items require a short one- or two-word answer or the filling in of a blank.

Examples

What do spittle and rubbish breed?

(Answer: Microbes)

Solve:
239
-143
_____
=

(Answer: 96)

The manometer of the sprayer shows that it
has _______________.

(Answer: Pressure)

Short answer and completion items have to be written carefully so that more than one interpretation of the question/incomplete sentence is not possible. The wording of the item should elicit the information specifically required.

Matching. Matching involves pairing of items from two different sets or columns because of their similarity or correspondence according to some rule or relationship.

Example

Column 1	Column 2
(1) Ecology	(A) The pattern of interconnected food chains
(2) Predation	(B) The taking in and using of organic food for energy, growth and replacing cells
(3) Nutrition	(C) The study of how living things relate to each other and to their nonliving environment
	(D) A relationship between two kinds of organism in which one benefits by killing and eating the other

Matching items should be kept relatively short. Note that there are three choices under Column 1 and four choices under Column 2. This insures that matching will involve deliberate choices in all cases under Column 1. If a choice under one of the columns is usable more than once, make that information available to learners as a part of the question.

Multiple-choice. Multiple-choice items are the most versatile and effective form of test items. A multiple-choice item has a stem, followed by multiple options from which one or more could be selected.

Example

A farmer should do early weeding of his cotton crop:
[Stem]
(a) So that the cotton is not choked
(b) So that weeds do not consume the plants' food
(c) So that cotton gets enough air
(d) So that cotton has access to light
(e) So that cotton gets enough water
(f) Because weeds could breed insects dangerous for cotton

(g) To allow better growth of cotton
(h) To get a good cotton yield

Please note that in this case most of the above options are correct. Choosing the right options and leaving out the incorrect ones will be like writing a short essay on the advantages of early weeding of the cotton crop.

Typically, multiple-choice items have no more than four or five options, unlike the item above which has eight options.

Essay. An evaluator developing an MIS would not typically write an essay type test, but it is not impossible to imagine. In a literacy class some essay or composition may be written by learners as part of their test at the end of a cycle of literacy instruction. This is the easiest type of test to write and the most difficult one to score. When essay questions are carefully written, specifying exactly what is required, essay questions do provide the students with opportunities to analyze, synthesize and evaluate subject matter content. Objectivity of scoring of essay type questions can be increased if teachers themselves write model answers to their own essay type questions and then judge student responses according to the model answers.

Simulations. Simulations of various kinds provide exciting teaching and testing possibilities. Various types of "In-Tray/Out-Tray" simulations can be designed to test the performance abilities of trainees in life-like decision-making situations.

Pre-testing tests for improvement

Good test items have to test what they are supposed to test; and should be well written so that they communicate the same meaning to all readers clearly and unambiguously.

Item writing takes time, patience and skills. With time and patience, skills can be developed. One thing that test writers must do is to pre-test their tests; and go through careful revisions of their tests on the basis of pre-testing.

After a more wide-scale use of a test in an evaluation study, the test should be revised once more. Even if you will never use it again, the revisions will train you to write better tests for future evaluation studies.

Time tests, power tests and other considerations in administering tests

Tests should be administered so as not to make respondents afraid and anxious -- what is called "test anxiety" can become a serious problem. Indeed, within developmental settings, where we deal with adults (and also with government functionaries), we may find that we want to give a test but the adults concerned do not want to take the test. Sometimes a few test items may have to be hidden in an opinion questionnaire or an interview schedule. When administering a test, the respondents should be comfortably seated and instructions in how to complete the test should be fully explained.

Finally, tests can be time tests or power tests. Time tests have to be completed within a particular period of time: 45 minutes or an hour, for instance. At the end of this time, test papers are collected whether or not these have been completed. Power tests are given to determine how much the respondents have learned (and not how fast they can answer questions). In a power test, there are many more test items than there are in a typical time test, and time is allotted generously to students for completion of the test.

SECTION D: Testing Attitudes, Observing Actions and Results

In real-life work settings, tests and checklists and simple surveys will be the most widely used instruments in the design of an MIS. The currently held or changing attitudes of learners and participants will be studied as part of full-fledged evaluations using the NE or the RE approaches. The same will be true of the studies of impact involving the practice of new skills and adoption of ideas and innovations.

MIS's may, however, include *some* attitudinal and observational data. In our example of an MIS for a functional literacy project, MIS data may include learners' answers to attitudinal questions such as these: Do you think schooling of girls is as important as schooling of boys? What use do you personally expect to make of your ability to read and write?

The literacy teacher (or another extension worker) may visit the adult learner's farm and home and observe adoption of innovations such as construction and use of garbage pit and latrine, planting cotton in rows, or use of fertilizers and pesticides, and can enter the observations on a checklist. With very little processing, weighting and coding, this information can be added to the MIS.

In this handbook, we shall discuss the use of unstructured interviews and observations in Part IV: Evaluation in the Naturalistic Mode. Questionnaires, attitudinal scales and structured observation will be discussed in Part V: Evaluation in the Rationalistic Mode.

SECTION E: Data Analysis

The use of registers, forms and tests in literacy classes and income-generating groups; and the application of other instruments in the community -- interview schedules, opinion surveys and observation schedules -- will give us lot of data items and data sets. These data will have to be *processed* and *analyzed* for us to be able to answer the questions initially asked in an evaluation study. Figure 5 on the next page illustrates this need diagrammatically.

The two tasks of "data processing" and "data analysis" overlap quite a bit. Data processing makes data ready for data analysis. Thus, data processing means the collation, consolidation, tabulation, and display of data in formats convenient for subsequent data analysis. Data analysis is the process of using a variety of logical, analogical, qualitative and quantitative operations to tease out of data answers to the questions initially asked in an evaluation study.

In the case of MIS data, data processing will typically mean tabulation of data. Again, in the context of MIS data, data analysis will typically mean visual analysis of data sets; and a few operations of descriptive statistics such as working out means and percentages and other ratios for use in bar graphs and pie charts.

Designing forms and tables

Designing forms, tables, registers, and formats for periodical reports has emerged as a speciality in its own right. We suggest that in designing such forms we should learn from other projects. If at all possible, obtain a set of forms available for similar projects in other departments (or countries) and try to adapt them for your own use. Always pre-test your own set of forms, tables, registers and reports before printing big orders.

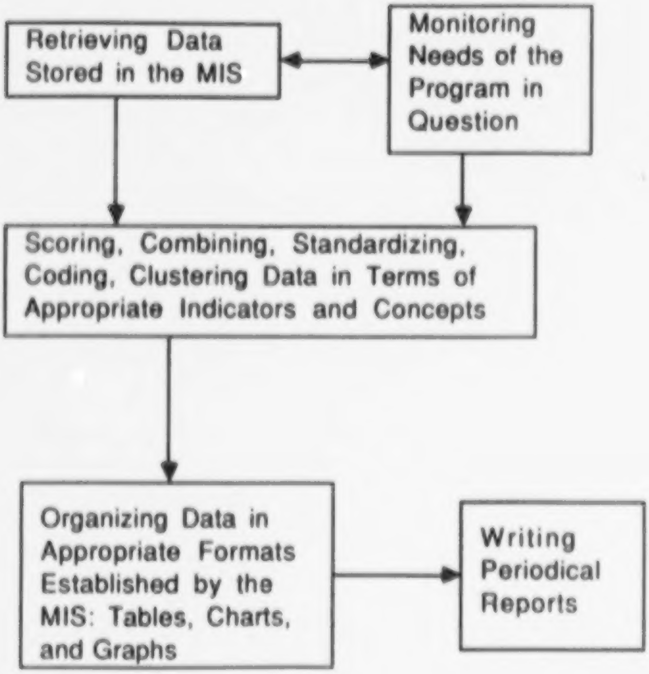

Retrieving Data Stored in the MIS
Monitoring Needs of the Program in Question
Scoring, Combining, Standardizing, Coding, Clustering Data in Terms of Appropriate Indicators and Concepts
Organizing Data in Appropriate Formats Established by the MIS: Tables, Charts, and Graphs
Writing Periodical Reports

Figure 5: MIS - Focus on Data Processing and Analysis

Graphic display of data

MIS data are most frequently processed for display in charts and graphs of various kinds. Line graphs, bar graphs and pie charts can be used quite effectively to present useful information to decision-makers. It is not within the scope of this handbook to teach the design and production of graphs and charts. Some references will most probably be available in local libraries. Statistical publications from the UN sources, the World Bank, national departments and agencies dealing with development and economics, and even periodical literature, will contain a variety of graphs and charts that should provide useful ideas for presentation of statistical data from the MIS.

Things to do or think about

1. Undertake the concept analysis of the concept "self-reliance", individually, on your own. Then, do another concept analysis of the same concept, participatively, in a group. Are there differences between the two versions of concept analysis?

2. Develop a set of indicators for an "effective literacy teacher" in your setting. How is it different from the example used in the handbook?

3. Develop a *detailed* list of knowledge items, principles, skills, and attitudes that you want your adult learners to have learned by the end of your literacy course. Write a test or set of tests using only the knowledge items.

4. Have you been asked questions recently by someone as part of an evaluation or a survey of some kind? What do you remember that was good about the interview? What did you find irritating or unacceptable? Was the interviewer able to win your trust?

5. What would you observe if you wanted to include one or two pieces of information on "Working Habits in the Office"?

6. Look at a table of literacy statistics used in one of the most recent reports from a literacy program or development agency.

Is there some useful information that is missing from this table but could easily have been included in the table? Does the table as designed help in the "visual" analysis of data presented?

Notes

1. Gooler, Dennis D. "The development and use of educational indicators" in *Educational indicators: Monitoring the state of education*. Proceedings of the 1975 ETS Invitational Conference. Princeton, N.J.: Educational Testing Service, 1975, page 15. See also Francette, S. *Indicator-based educational and cultural classification, grouping, and statistical analysis of the 25 least developed countries*. Paris: Unesco, Division of Policy and Planning, 1977.

2. Bloom, B.S. et al. *Taxonomy of educational objectives, Handbook I: Cognitive domain*. New York, NY: David McKay, 1956.

3. Krathwohl, David R. et al. *Taxonomy of educational objectives, Handbook II: Affective domain*. New York, NY: David McKay, 1964.

CHAPTER 8

WRITING PERIODICAL AND SPECIAL REPORTS BASED ON MIS DATA

Periodical and special reports must be written on the basis of MIS data and made available for use by program managers at the various levels of the literacy system. Periodical reports will use data already available in the MIS and present it in an agreed form to managers at agreed intervals of time. Special reports may make a more exhaustive use of the data in the MIS, or may involve collecting additional data specifically for the purpose of the special report. To be useful, periodical reports must present national data on the context of the program; a profile of the program describing program development over time; and information that catches the dynamics of the program, making useful comparisons and showing interesting connections.

Useful data are generated by literacy campaigns, programs and projects in the very process of their implementation. Of course, these data do get used by decision-makers in some ways in their day-to-day work. But such use is often reactive, impressionistic, unconnected and unsystematic. In this handbook, we have suggested that data generated by programs in their implementation be collected systematically and comprehensively and then used pro-actively in decision-making. Unfortunately, not all literacy initiatives install MIS's.

Even more unfortunately, and inexcusably, some programs have established MIS's, but do not use them well to develop useful information for their day-to-day decisions. In such cases, some sort of data keep on flowing up to the Headquarters but no one does anything with it. Decision-makers have not learned to use "information" in the management of programs. They do not miss the information that a well-functioning MIS could have provided them. In turn, they do not make available to their organizations the minimum of resources of time and personnel necessary for writing reports based on MIS's. It is indeed a vicious circle.

Our discussion on writing periodical and special reports is done from the vantage point of the total system level. We will be talking here of reports which will be written for and by the top managers, presenting a profile of the total system for making system level decisions. We must, however, keep in mind the fact that any system level report is a culmination through consolidation of a multiplicity of reports generated at lower levels of the system from adult literacy groups in villages, to counties, districts or provinces, and to zonal or regional levels.

It should also be stated that our discussion in this chapter is restricted to more or less standardized periodical reports. Special reports that make use of MIS data combined with data specially collected in the context of an RE study will be discussed separately in Chapter 16.

An ideal report is the one that makes the best use of the data in the MIS; presents a picture of the program that can be easily understood; presents information in a form that is suitable for use in monitoring and decision-making; and last but not least, is timely.

In an earlier chapter we also talked of the necessary and sufficient dimensions of an MIS. We can now say that a periodical report should present at least three types of information:

1. Information on program context -- which need not be collected within the program by the MIS functionaries but may indeed be obtained from other relevant government sources such as the census bureau or the offices dealing with economic, social and educational statistics
2. Information on program status which describes in a few well-designed tables the current status of the program, and particularly the structure of access and achievement, and
3. Information on program dynamics showing before and after trends and connections between various aspects of the program

The title page

The periodical report should be properly and accurately titled. It should be possible quickly to find out the period to which the report pertains. The date when the report was actually completed for submission to higher authorities should also be indicated. In other words, the lapse of time between the program period and the completion of the report should be easy to find.

The volume and number of the report should also be indicated. Is it the first year ever that such reports are being published? Is this the first report of the year? If it is the second of the four quarterly reports published every year, and if it is the fifth year of such reports, the report should be marked Volume 5, No. 2.

Context of the report

The literacy program should be placed within the context of national development in the country. The role assigned to literacy within the on-going social and educational change should be briefly stated. If at all possible, figures of literacy at the national level should be provided to enable readers to understand the size and scope of the particular literacy program being reported upon.

Preface: Changes in boundaries, categories, definitions and indicators

The report should include a short preface to warn readers about conceptual and definitional changes in the program since the last report. Sometimes changes can be made in the boundaries of provinces and districts that may increase or decrease figures of enrollment and dropouts. The program may decide to categorize dropouts differently from before, affecting data in the MIS during that particular period. Definitions of literacy may be adjusted upwards or downwards with concurrent changes in tests of literacy and numeracy. This, again, may change figures. Indicators may be changed as well: instead of administering attitudinal tests, the program may decide to go to more easily obtained self-reported data. This may change the complexion of figures in the MIS. Whenever such changes are made in the program, the reader of the report should be suitably warned so that the report can be properly read.

Main introduction and sectional introductions

The report should begin with an abstract-like statement that describes the salient features of the report as a whole. Thereafter, each section of the report should itself begin with an introduction.

Data shown in the form of tables

The data should be tabulated, as in Table 1 on page 146, to show the changes in the structure of the program along the time dimension. Levels of participation and achievement, and such other factors as training and experience of teachers can also be expressed conveniently in tabular form, as in Tables 2-8 on pages 146-150.[1]

TABLE 1
DATA ON THE GENERAL CONTEXT OF THE PROGRAM

Year	Total Learners Enrolled	Male	Female	Male/Female Ratio
1988				
Urban				
Rural				
Total				
1989				
Urban				
Rural				
Total				
1990				
Urban				
Rural				
Total				

TABLE 2
PARTICIPATION OF MEN VERSUS WOMEN AS FUNCTIONARIES

Category	Male	Female	Male/Female Ratio
Program specialists			
Supervisors			
Teachers			
Learners			

TABLE 3
COMPARISONS WITH THE REPORTING PERIOD IMMEDIATELY PRECEDING

	Last Quarter (October-December 1989)				Present Quarter (January-March 1990)			
	1	2	3	4	1	2	3	4
Region A								
M								
F								
Total								
Region B								
M								
F								
Total								
...								
...								
...								
Region X								
M								
F								
Total								

Key
1 is the number of enrollment figures brought forward
2 is the number of new enrollees
3 is the number of those who dropped out
4 is the number of those who graduated from the program

TABLE 4
PERFORMANCE IN LITERACY (R/W) AND NUMERACY (N) LEARNING, FUNCTIONALITY (FN), AND AWARENESS (AW)

Number	Score Structure											
	R/W			N			FN			AW		
	F/3	S/3	T/3	F/3	S/3	T/3	F/3	S/3	T/3	F/3	S/3	T/3
Male												
15-25												
26-35												
36-45												
45+												
Female												
...												
...												

Key
F/3 stands for top third,
S/3 stands for second third, and
T/3 stands for the lowest third.

TABLE 5
TIME TAKEN TO COMPLETE DIFFERENT STAGES OF LITERACY ACHIEVEMENT (1989 COHORT, ALL REGIONS)

Years in the program	For completing					
	S1		S2		S3	
	M	F	M	F	M	F
0-1 years						
1-2 years						
2-3 years						

Key
S stands for stage of learning completed.

TABLE 6
PARTICIPATION IN INCOME-GENERATING ACTIVITIES

No.	Project Title	Year Initiated	Region/ Location	Coverage	
				M	F
Personal benefit orientation					
1.					
2.					
3.					
...					
Community benefit orientation					
...					
X					

TABLE 7
TRAINING OF TEACHERS

	Total (BF)	Left	New Recruitment	Trained
1985				
1987				
1989				
...				

TABLE 8
AGE, EDUCATION, EMPLOYMENT AND OCCUPATIONAL INTERRELATIONS BETWEEN MALE AND FEMALE TEACHERS

	Number	Education	Years in employment	Previous/ concurrent occupation
Male				
15-25				
26-35				
36-45				
46+				
Females				
15-25				
26-35				
36-45				
46+				

Data about post-literacy and learner satisfaction with the program

Comparable tables can easily be devised for post-literacy programs, to store information on learner satisfaction and subsequent (post-literacy) achievement.

Simple correlations based on MIS data

We have not discussed statistics for working out comparisons and correlations in this Part of the handbook. However, MIS data can be used to work out information on the following:

- Reading and functionality correlation
- Reading and awareness correlation
- Numeracy and functionality correlation
- Reading and satisfaction correlation
- Gender-based comparisons

- Class-based comparisons, and
- Regional comparisons

Data display by tabulation

A useful report based on MIS data will always include tables and sometimes bar graphs and pie charts. Tables, if they are to be useful, must be accurately compiled and should be easy to read and interpret. In the local library or in a bookstore, you may be able to find a manual for writers of term papers, theses and dissertations. Such manuals provide excellent help on how to compose tables. It is not within the scope of this monograph to provide detailed instructions on how to make tables. We shall be satisfied with making the following general suggestions:

1. Number your tables as TABLE 1; TABLE 2; etc.
2. Give a title to each table; and make the title both accurate and complete.
3. The headings and descriptions used for rows and columns should also be accurate and complete.
4. Use correct placing and spacing, especially where numbers and decimals are involved.
5. Do not make up your own abbreviations. Use only standard abbreviations. Even when standard abbreviations are used in a table, explain them in the footnotes to the table.
6. Sometimes, statistics from different years may have to be used in the same table. Indicate which year those statistics belong to, e.g.:

Population figures in millions (1975)	Per capita income (1974)	Radio sets in use (1975)

7. Separate "estimates" from "factual counts". Do not confuse one with the other.

8. Wherever necessary, qualify your data. For example, you may have to say: Figures do not include data from Korea; or Domestic workers have not been included, etc.
9. Standardize your scores, if at all possible. However, if standardized scores are misleading, also include absolute scores.
10. Sometimes, comparative statistics may have to be included in tables to make sense out of a given set of statistics. One can get a better idea of the level of poverty in a country by seeing, in the same table, the per capita income figures from the U.S.A. or Sweden, or even from a richer neighboring country.

Data display by graphics

The question of preparing graphics for displays of data is important. Graphics communicate ideas simply and attractively, but they are not always easy to make. There is a lot to learn about making graphics. It may interest readers that there is a special national Council on Social Graphics in the Bureau of Social Science Research in Washington, D.C., which recently held a general conference on the topic of "Graphics for Data Analysis and Social Reporting". It is not within the scope of the present monograph to discuss the preparation of graphics for data display at any great length. See any standard manual on making graphs and graphic displays in the local library.

Note on special reports

Special use of MIS data may be made by evaluators using either the NE or the RE approaches. Evaluators in the NE mode will make descriptive use of data in the MIS. Evaluators using RE can take samples from data already in the MIS data base and test different types of assertion.

Note

1. The dummy tables presented here were actually used by the author in conducting an evaluation of a national program implemented by a voluntary association in Africa.

Part IV

Evaluation in the Naturalistic Mode

Naturalistic Evaluation (NE) was first introduced in Part I, Chapter 2, of this handbook. Another brief description of NE was included in Part II, Chapter 4 where NE was presented as one of the three components of the methodological triangle of evaluation: MIS, NE and RE. In the general introduction to Part III, we suggested that the MIS should be considered to be the most important, indeed, the necessary part of the methodological triangle of evaluation; and that MIS should get first priority in the allocation of resources.

What we shall now suggest is that the second priority in the allocation of resources should typically go to discovering *qualitative* changes in the lives of individuals and communities through NE. These are not recommendations emerging from some methodological dogmatism. In fact, these recommendations have arisen from long and varied experience in conducting and teaching evaluation in the Third World settings. It is a matter of fact that MIS data and NE data have been found to be the most widely used data by decision-makers in their day-to-day decisions.

Our discussion of NE will be divided into the following chapters:

9. Naturalistic Evaluation -- Theory, Questions and Design
10. Writing a Proposal for an Evaluation Study in the Naturalistic Mode
11. The Process at a Glance: Tools and Techniques of Naturalistic Evaluation, and
12. Writing Reports on Naturalistic Evaluations, and Writing Periodical Reports Naturalistically.

CHAPTER 9

NATURALISTIC EVALUATION -- THEORY, QUESTIONS AND DESIGN

Naturalistic evaluation (NE) seeks to study reality naturally -- as a whole, in all its complexity; in its own particular context; in its perpetual flux, without trying to simplify and reduce it to a manageable evaluation design. The goal of design in an NE study is to ensure trustworthiness, which in turn depends upon credibility, fittingness, auditability, and confirmability of the study.

The discussion of the Naturalistic Paradigm in Part I, Chapter 2, should provide a conceptual umbrella for a discussion of the theory of NE in the present chapter. A few remarks are included below by way of recollection and further explanation.

NE theory and methodology

The theoretical and methodological issues of evaluation, we have suggested earlier, can be summed up in the form of two interrelated questions: What is the nature of reality? and How should we go about making knowledge-full and informative assertions about that reality?

Of course, NE is rooted in a particular theory of "reality" and our "knowledge" about that reality. According to the naturalistic paradigm, all reality is not "out there" for everyone to see and record. Reality is a "social construction". In other words, as individuals, we construct our own individual realities; and we all carry our own special meanings about the world inside ourselves. Not that for the five billion or more people alive today, there are five billion *or more absolutely* different realities! Social interactions within families, communities and cultures do create realities shared at various levels of commonality. A large part of our world is thus already constructed for us.

Yet, within the *shared* commonalities of communities and cultures, there are realities that are *unique* to individuals. These unique versions of reality, these meanings, are often so important that they must be studied as uniquely held by individuals or groups, and not be lost in our attempts to make universal laws about human nature.

This brings us to the conception of the nature of knowledge in the naturalistic paradigm. Knowledge, in this paradigm, is not universal: part of it may be quite general, and a part of it is particular. Knowledge is contextual, though context may vary in its scope and its temporal life. Finally, knowledge is rooted in history: it is not good and true for all times. In NE, therefore, we talk not of "generalizations" but of "insights" for transfer to other settings and times.

In NE, the methodology for studying reality should be holistic. The real world should not be factored and fragmented to test hypotheses, to study causalities, and to make predictions. In this systemic and dialectical world, we should be looking not for "networks of causalities" but for "networks of plausibilities", and instead of seeking to predict, we should aspire to building reasonable expectations.

Since reality is a social construction, products of knowledge produced by the evaluator will also be individual constructions. It will be absurd to apply to such knowledge statements, the rationalistic criteria of reliability and validity. Of course, we shall have to apply some set of criteria to give our evaluative statements the status of "warranted assertions". We shall talk about these criteria later in this chapter.

The naturalistic evaluator uses "self" as instrument and thereby accepts the subjective nature of all evaluation and research. There is a unity between the knower and the known. What the evaluator offers is a social construction that has been built on the basis of a "sharp intellect" and a "clear perception" and refined within the questioning dynamics of participation and collaboration with others. Thereby, the idiosyncratic subjectivism is taken out; a multiplicity of realities is often presented rather than one single truth; and an overall statement about reality can be made that "holds" as "objective" reality in that context at that time. We are once again back to the concept of "warranted assertions", based on data that are vivid, useful and credible.

Considerable work has since been done in the area of NE as definitions have been proposed, design issues discussed, and methodologies elaborated.[1]

A frequent confusion has arisen as "techniques" of data collection have been equated with "methodology". It has to be understood, for instance, that the use of qualitative techniques of data collection does not make an evaluation study a study in the NE mode. Indeed, both RE and NE do make use of qualitative techniques. What is crucial are the ontological assumptions (about the nature of reality) and the epistemological assumptions (about the nature of knowledge) that are made and how the data are processed, once data have been collected. Thus, ethnographic techniques in the collection of data would not ensure that NE approaches were being followed if the data so collected were later fitted into nominal or ordinal scales and were statistically treated.

There is, however, one evaluation approach that is quite congenial to the NE methods: Participative (or Participatory) Evaluation.

NE and participatory evaluation.

Participatory evaluation, as the name suggests, is conducted in participation with the people and publics concerned. Evaluation becomes both educational and liberating. Essentially, participative (or participatory) evaluation is one that is conducted in mutual collaboration by all those engaged in the conduct of a program. At its best, the organizers play a facilitative role while the people being served by particular programs take charge. These participants determine, through "dialogic action", what the evaluation needs are, what information should be collected and how, and what norms and standards should be used to judge success or failure. Participative evaluation has the same assumptions as does NE, with the added feature of strong ideological commitment to the cause of those being served by programs.[2]

Negotiation and collaboration

NE, as we have defined it in this monograph, includes both negotiation and collaboration. Reason and Rowan, in the volume they edited on new research paradigms, published in 1981, consider *collaborative* inquiry as being the essence of human inquiry.[3] In their 1989 book, Guba and Lincoln have called ***negotiation*** in its broadest sense the key dynamic of what they term "Fourth Generation Evaluation".[4] It is through negotiation that the evaluator is able to empower people and join knowledge with action.

The reality context of NE

Cronbach[5] has made a distinction between two contexts of reality: the context of accommodation and the context of control. In the context of control, rationalistic approaches are possible. But in the context of accommodation, we have no choice but to follow naturalistic inquiry approaches. Two points should be made here: One, that a naturalistic evaluator without betraying himself or anybody else, will sometimes be doing studies that follow rationalistic evaluation assumptions and methods; and, two, that naturalistic evaluators, in the context of NE, will continue doing a lot of counting and measuring.

Questions for NE -- in the context of accommodation

The scope of NE is wide indeed. NE can seek and find more meaningful answers to *most* of the questions we have listed earlier in Chapter 5 as suitable for MIS. And NE can answer *most* of the questions we will later list in Chapter 13 as questions suitable for RE. NE does, however, have its favorites. It is best suited to answer questions about *qualities* of inputs and outputs, about the *nature* of processes, about human experiences with curricula and programs and about program impacts on groups, institutions and communities, and about the *totality* of contexts which simply do not fit either the data sets of the MIS or the research designs of RE.

NE is perhaps the only way to go when we do not even know what to look for and what questions to ask of people. As Frederick Erickson[6] put it insightfully and succinctly, the essential question in NE is: What is happening? This is a perennial question and a very significant one. From it one should be able to see that in NE a question is not asked as in the context of a test, or in a structured interview so as to obtain one correct answer to the question. Questions in NE are excuses for starting long and rambling *conversations* between evaluators and respondents wherein several unanticipated questions are raised and many meaningful answers are constructed.

In the more concrete context of the evaluation of literacy and post-literacy campaigns, programs and projects, the following sets of questions can be raised. The list, of course, is not exhaustive, but merely illustrative:

1. How do illiterate individuals, men and women, in various stations of life, explain their present condition of illiteracy and "disadvantage"? Do they indeed see illiteracy as a disadvantage? What is their "mythologic" that explains their suffering?
2. How do illiterates, men and women, farmers and workers, in cities and villages, survive in cultures built on the assumptions of print? Those who have become semi-literate, what kinds of symbioses have they built between literacy and oracy?
3. What are the needs of the illiterates as they see them and what are their expectations from the program being offered? What was going through their minds when they first joined the program?
4. How are they experiencing the program? Is it useful? Is it inconvenient? Is it contributing to their self-esteem? What is it doing to their identities? Is somebody listening to them or is it one more intervention in their lives that they cannot fight to keep out?
5. What do program participants think are the purposes of policy-makers? Would they rather have programs run by local NGO's or the church, mosque, wat, mandir or gurudwara?
6. What part of the curriculum do they like best? Would they rather learn skills now and learn literacy later? Would they rather learn to empower themselves in relation to the local leadership, the politicians and the state rather than learning the 3 R's?
7. Which of the teachers teaching in the community do they like and why? What expectations do they have of their teachers?
8. What do they think of being tested in reading, writing and numeracy? Do they think tests should be in class or should they be national tests as in Tanzania, where tests are held once every two or three years on an appointed day, so that those who want to do so may take them?
9. Is literacy usable once it has been acquired? In what ways?
10. What do they think of the language of literacy? Should literacy have been taught in the mother tongue or in the

national language? Do they want to learn literacy in English? or French? Why?

11. Was their participation in the program worth the time?
12. What would they like to do with their literacy skills? What kinds of post-literacy programs would they want to have?
13. What books have they read recently? What did they like about them? Would they themselves like to write something for other new literates to read?
14. Is there something happening to the community as a whole? Any changes in social, economic and political relationships? Are leadership patterns changing?
15. How has the village library or an institution like the Folk Development College changed realities around them?
16. How has literacy affected radio listening in the community?

These are the questions that can be asked of learners. Similar questions can be asked of spouses and children of new learners, of teachers, local leaders, district officials and so on. Answers to these questions can be most illuminating.

By way of summarizing, the following conceptually relevant points should be made about NE:

NE searches for meanings: NE is not interested in behavior (the physical act) but in actions (the physical act plus the meanings held by those involved in the act). Thus, its credo is: meanings-in-actions. First, emic (insider's) meanings are developed, and that means that a multiplicity of meanings are delineated. Then, etic meanings (meanings as objectified within some collectivity) are delineated.

NE recognizes multiple causalities, not linear causal links: With the concept of multiple realities comes the concept of multiple causalities. Cause is not linear, cause is multidirectional: social entities are in a continuous process of mutual definition of each other.

NE addresses multiple layers of universality: To quote Erickson, the naturalistic evaluator is interested "in particularizability rather than generalizability. One discovers universals as manifested concretely and specifically, not in abstraction and generality." There are different layers to universality. The innermost core may be specific to the life of a particular group at a particular time, while outer layers may apply successively to other programs, other settings, and even other cultures and times.

NE does not test theory, it uses grounded theory: NE does not test hypotheses generated by theory and then return to theory to enrich it. It uses grounded theory and then expands theoretical understandings.

NE theory and methodology have also been made concrete in terms of the design of NE studies and in the special methodology of NE. To these we turn now.

Design in naturalistic evaluation

One reason for the immense popularity of rationalistic evaluation (RE), as we shall see, has been its concreteness. The RE paradigm is able to suggest internal and external validity, reliability, and objectivity as the pillars of all evaluation design. It is then able to suggest standard experimental or quasi-experimental designs, sampling procedures, assignment of subjects to treatments, experimental controls, methods of instrumentation, and statistical formulas that will protect the validity, reliability and objectivity presumed to have been initially established.

NE does not accept the assumptions that RE makes about our world. It is a great frustration, however, that NE continues to be tested by the naive against the norms of RE. Questions keep on being hurled at naturalistic evaluators about reliability and validity, and then about the objectivity and generalizability of their conclusions.

Some of the problems of NE may be of its own making. It is only recently that any concretization of NE procedures is taking place either in design or methodology. We realize that assertions such as "NE design is 'emergent' design"; and "NE methodology uses the 'human instrument' as the instrument of data collection" are not enough. The work of Guba and Lincoln, already referred to above, is beginning to provide concrete procedures and tactics whereby the evaluator could be ready for the design to emerge (without being outwitted and perplexed), and whereby he or she could be "systematic" and "objective" in the study of subjective reality.

The design in the evaluation design

The design (the hidden purpose) in any evaluation design is to meet certain criteria in methodology that will ensure that the best "truth statements" can be made about a particular reality. In RE, as we suggested earlier, they are internal and external validity, reliability and objectivity. In NE, the criteria to be met are those of trustworthiness, which in itself is made up of four components: Credibility (internal validity in RE), Transferability (external validity in RE), Dependability (reliability or replicability in RE) and Confirmability (Objectivity in RE).[7]

Credibility

Credibility is ensured through prolonged engagement with the people in a program and persistent observation in the field. You stay in the field long enough, and spend enough time with your respondents. Credibility is also built through triangulation of sources of data, of methods used, and of investigators or investigator teams. Peer debriefing (exposing oneself to a disinterested professional peer) is likely to keep the evaluator honest and alert and will, therefore, increase the credibility of the evaluation product. Member check, that is, soliciting reactions to your findings from the respondents whose realities are being described by the evaluator, contributes to credibility. Finally, negative case analysis helps. One must make an assiduous search for the negative cases that seem to go against the general understanding, and appropriately qualify all assertions.

Transferability

Thick description of context and pattern will help transferability. People will be able to hear in the descriptions echoes of their own realities, and be able to receive not instructions but useful *insights* -- generalizations rich with particulars.

Dependability and confirmability

Use of audit of both the process and the product will contribute to dependability and confirmability.

The advice to the designer of an NE study is to do whatever can be done to increase credibility, transferability, dependability and confirmability. But that may not be enough. What does it mean to go to the field without *a priori* theory, looking for meanings, and so on? We are going to suggest that the naturalistic evaluator go to the field not without theory, which is impossible anyway, but go there with an "empty conceptual set" that will enable him or her "to inherit the wealth of knowledge of social sciences, without the conditions and categories imposed by the trustees of this knowledge". Since evaluation is primarily integrated with change, it would be better if the empty conceptual set we are talking about were subsumed under a change model. Such a model is available as Bhola's CLER model.[8]

Using the CLER model of change as an outline for flexible design

It is impossible to go into the field with an *absolutely* open mind and to let theories come from the ground, designs emerge, and themes shout themselves out. The real intention here is that we go to the field not in the experimental mode but in the dialectical mode. We do not go as theoretical orphans in the tradition of social sciences, but we do not go to test theories either. We look at the reality and watch for the patterns in which it seems to be embedded. We do not have hypotheses to test, but we do have hunches; and we want to be careful that even these hunches do not impose selective perception on us, making us miss what is really going on. We do not have random or statistically selected samples, but we do have ideas about the sources of data, the most useful places to begin before the snowballing process takes over. We do not have structured and pre-tested instruments, but we do have ideas about the themes on which questions will be asked, and how the interaction between the evaluator and the subject might be managed. Finally, we do not search for universal laws, but we do want to make statements about reality that are true in the context and can generate insights (not instructions) for programs and practitioners elsewhere.

What one should, indeed has to do, is to go into the field with a model so flexible that it can serve as an "empty conceptual set of containers" into which one can then collect the realities as one finds them. Since evaluation seeks to measure *change* in the performance of program systems, a model that deals with change would be preferable. A model that meets both these criteria (deals with change and is an empty conceptual set) is provided by the CLER model, to which we briefly return.

C in this model stands for configurations and configurational relationships. L stands for linkages between the planner system and the adopter system and linkages among the actors in both these systems. E is the environment surrounding the planning system and the adopter system -- and the environment surrounding these two may not be the same. R stands for resources -- the planner system needs these to promote change, and the adopter system needs them to incorporate change.

The model suggests that to promote change, one should, synergetically, optimize all the four variables. To evaluate change, one should see "what is happening" to the meanings held by configurations and to the quality of relationships among and between them; what is happening to the linkages among people, groups, and institutions; what is happening to the generation and allocation of resources; and what is happening to the environment within which individuals and groups and communities are living. In other words, one should ask all those questions which Erickson suggested are the special preserve of interpretative research and evaluation.

Thus, one needs to go into the field and see the changes in the C, L, E, R ensemble at various points in time:

Time (1) Descriptions in terms of C, L, E, R
Time (2) Descriptions in terms of C, L, E, R
....
....
Time (n) Descriptions in terms of C, L, E, R

Some further discussion of the topic above is included in the next chapter, which deals with the topic of writing a proposal for an evaluation study in the NE mode.

Some popular NE designs

Following the example of RE, which has a whole series of fixed experimental and quasi-experimental designs, in NE we are seeing the beginnings of "ways of going about it" that can be shared with others as procedures and patterns. Some of these ways of going about NE may turn out to be designs of sorts.

Opinion survey through hermeneutic circles

What are called hermeneutic circles may someday be seen as the NE answer to the RE's opinion survey design.

In the School of Education at Indiana University, there was the need to develop a statement that would reflect a faculty view of the future of the School of Education until the Year 2000. Instead of a typical survey, a NE was conducted, using the hermeneutic circles as a "design". Twenty faculty members volunteered to serve as group leaders. Each chose to talk to five faculty members. The group leader (G1) talked to one faculty member (F1), first. The constructions from this G1-F1 encounter were presented to faculty member (F2) and so on, until all the five members had been covered. There was supposed to be a second iteration, but the exigencies of time made that impossible. At the end of the first iteration, the group leader (G1) met with all others (F1, F2, F3, F4, and F5) as a group. All the 20 such groups went through such a process.

The 20 group leaders (G's) divided themselves into four groups of five each and coopted, in each group, 2 additional members, all from the Long-Range Planning Committee of the School of Education that had commissioned this study. A second generation of constructions was thus developed.

In a third stage, twenty representatives (representing diversity of views, rather than diverse departments) were selected to sit on a committee which developed a report. This report did not attempt to develop a consensus but portrayed all the divergences that had been met.

This final report and recommendations were sent to all faculty to "vote" their acceptance of the full report, or the report in part, with qualifications as necessary. All the various reports and data were then given to decision-makers for their implementation.

In the opinion of the present writer this set of procedures qualifies as a "design" for opinion surveys. Other designs can be and should be built that are, again, enabling designs for those conducting NE.

Sampling as an element of design in NE

As we can see from the above discussion, design elements are built into the choice of samples, instrumentation and methods of data collection. These issues will be treated in detail in a separate chapter below. For the present, let us only remind readers that sampling in NE is not random, but purposive and what Guba and Lincoln have called serial and contingent.

We also want to warn naturalistic evaluators against developing an orthodoxy of their own. The unstructured interview and observations (with vignettes, anecdotes and direct quotes from the actors involved) will remain the naturalistic evaluator's most favorite techniques of data collection and presentation. It should be remembered, however, that NE will be using both induction and deduction; narrative on the one hand, and analytic charts, summary tables, and descriptive statistics, on the other.

Things to do or think about

1. In your cultural tradition, is there something which reminds you of the theory and methodology of NE?

2. Are you personally convinced that NE is the way to go in your particular program? How could it help? Can you convince the authorities above you to support NE? How? With what expectation of success?

Notes

1. See Lincoln, Y.S. and Guba, Egon G., *Naturalistic inquiry*. Beverly Hills, CA: Sage, 1985. Also, Williams, David D. (ed.), *Naturalistic evaluation*, special issue of *New directions for program evaluation*, No. 30, June 1986.

2. For a definition and discussion of dialogic action, see Freire, Paulo, *Pedagogy of the oppressed*. New York, NY: Herder and Herder, 1970. For participatory evaluation refer to the work being done under the aegis of the International Council for Adult Education, Toronto, Canada.

3. Reason, P. and Rowan, J. (eds.) *Human inquiry: A sourcebook of new paradigm research*. New York, NY: John Wiley and Sons, 1981.

4. Guba, Egon G. and Lincoln, Yvonna S. *Fourth Generation Evaluation*. Newbury Park, CA: Sage, 1989.

5. Cronbach, Lee, et al. *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass, 1982.

6. Erickson, Frederick, "Qualitative research in teaching". In M.C. Wittrock (ed.), *Handbook of research in teaching* (3rd. edition). New York, NY: Macmillan, 1986.

7. Guba and Lincoln: see Note 4 above.

8. Bhola, H.S. "Planning change in education and development: The CLER model." *Viewpoints in teaching and learning*, Vol. 58, No. 4, Fall 1982, pp. 1-35. Also, Bhola, H.S., "The CLER Model of innovation diffusion, planned change, and development: A conceptual update and applications." *Knowledge in Society: An International Journal of Knowledge Transfer*, Vol. 1, No. 4, pp. 56-66, Winter 1988-89.

CHAPTER 10

WRITING A PROPOSAL FOR AN EVALUATION STUDY IN THE NATURALISTIC MODE

The use of emergent designs and the human instrument in Naturalistic Evaluation (NE) does not mean that it can be conducted without any formal preparation; and, therefore, the writing of a formal proposal is unnecessary. NE does indeed require thoughtful preparation as does any other type of evaluation or research. Proposals must be written to demonstrate that the evaluator is knowledgeable about the general context of the study; is well grounded in social science research and has some ideas of the themes to be pursued; has preliminary ideas about the sites and sources of data; has given thought to the possible instruments and equipment that might be used for recording data; has planned the logistics of data collection, data collation, and interpretation of data; has taken care to establish procedures for auditing of evaluation procedures and results; and is sensitive to the obligations to make reports to various stakeholders including the respondents in the study.

NE is not for weak and fuzzy minds that cannot handle the real stuff! Indeed, NE may be more demanding and more challenging than evaluation and research in the rationalistic mode that presents us with a world of certainties, with conceptual road maps clearly marked; and procedural steps and formulas for everything the evaluator is likely to come across.

NE methodology is difficult because it does not offer formulas but frames for thought and action. It never allows the evaluator to dispense with thinking and simply to follow instructions. It demands from the evaluator that he or she should see both the overall pattern and the specific detail; should both see and see through; and use, at the same time, the two great human inheritances of *logical thought* and *keen perception.*

NE is not the second best alternative for some evaluation questions. In Chapter 9, we listed the general types of question which Erickson[1] had suggested were in the domain of NE; and these were the types of question that indeed cannot be answered by evaluation in the rationalistic mode, without chopping them into parts, and thereby changing the phenomena under study. Let us recollect some aspects of adult literacy that could be evaluated in the naturalistic mode:

1. What has happened and is happening to a community as functional literacy classes for men and women come into their community and their lives?
2. How are participatory planning and participatory evaluation experienced by everyone involved?
3. What happens to adult literacy policy as it moves from the center through the provinces to districts and to development blocks?
4. Why, in some communities in Africa, do more men than women terminate their participation in the literacy programs? Are participation patterns for men and women in formal primary education, in other development programs, in church and in related social groups, in any way similar?

As we can surmise, NE is apparently more promising for conducting exploratory studies, and needs assessments; for conducting base-line studies; for problem definition and for inventing local solutions; for organizational research; and, finally, for the conduct of impact studies. Naturalistic approaches can also contribute useful data to other evaluation concerns such as personnel evaluation and curriculum evaluation.

Proposals for each of the various types of evaluation study will be somewhat different, but the following concerns are reflected in a typical proposal:[2]

1. The evaluation should enable the evaluator to develop a conceptual scenario within which the evaluation design can emerge, the samples can be developed, interviews and observations made, data collected and interpreted, and audit trails left behind, without too many unpleasant surprises, rude shocks or serious breakdowns.
2. It should enable the evaluator to develop proper logistics in regard to development of evaluation teams, training of evaluators and their assistants, and travel to and living arrangements at the field-work sites.
3. It should be an instrument of communication with others who may provide professional help, formal approvals, and budgets.

Elements in a proposal for NE

A proposal for an NE study must include the following aspects. We take the example of an impact study which is likely to include all types of question that must be raised and answered.

1. *The delineation of contexts*

The many layers of context surrounding the situation should be described briefly. It may include the larger developmental context of the country, but will certainly include the "means x ends" interactions built into the program. It will, of course, include a short history of the program or project in question with a thumbnail sketch of the present state of affairs.

One should note that this will require a study of the documentation of development policy and plans as well as statistics (from the MIS perhaps) on the program itself.

2. *The contexture of problems and the initial focus of interest*

What are the interlinked general problems, contradictions, frustrations, disagreements, and breakdowns that people in the program seem to experience? It is possible to organize all or some of these under an initial theme of interest that could be used as a lever to enter and open up the reality. The theme is then not the statement of the problem in the classical sense, nor is it a question. The theme should be used within a particular temporal-spatial context, and in a way that allows the emergence of both the problem and the solution, both the question and the answer.

3. *Illuminating the theme*

Naturalistic evaluators do not have to be intellectual orphans, unaware of the social scientific knowledge that might bear on their thematic interests. While they may not prepare a literature review in the traditional sense, they will go to literature that illuminates their contexture of problems. Much of this will be descriptive case study material from other similar programs and projects. This will, additionally, help the naturalistic evaluator to develop a comparative perspective on the project being evaluated, as well as create some predispositions. It would be useful, if not necessary, for the naturalistic evaluator to go to talk with experienced colleagues who may share with him or her their tacit knowledge and contribute to the evaluator's fund of "what to expect".

4. *A "frame on the flux" of field realities*

It is often said that the naturalistic evaluator does not go to the field to test hypotheses generated from *a priori* theory. It is also said that the naturalistic evaluator works with grounded theory. Neither of these statements should, however, be taken in an absolute sense.

The naturalistic evaluator may not have hypotheses, but he or she does have both an interest in some general themes and a bundle of hunches and conjectures. Again, it is impossible to leave all our theoretical baggage home as we go to the field. What we need as naturalistic evaluators are particular kinds of theory that are not in control but are in collusion with us in our search. To borrow terms from the literature on organizations and institutions, these have to be "enabling" or "convivial" theories.

The CLER model, as we suggested in Chapter 9, can provide such a model, which should be used at the stage of proposal writing to generate general scenarios. In a study of impact of a program, the CLER model could be used to *unpack* themes as follows:

Configurations
What configurations should be studied: Individuals, Groups, Institutions or Communities? What configurational relationships should be given special attention: Men-Women; Chief-Community; Cooperative-Individual Farmers? What aspects of their relationships should be studied?

Linkages
The extent, direction and quality of linkages among people determines the quality of life in a society. What is the class structure? Are leadership and institutions responsive to the needs of the people?

Resources
The CLER model talks of six types of resources: knowledge, influence, materials, personnel, institutions and time. Are appropriate resources available? Are resources well used? Is the community

generating resources of its own? Is a particular group or class capturing resources that should justly go elsewhere?

Environment
Is there an environment of hope? What are the signs? Are hopes justified?

As we can surmise, this examination in terms of CLER will provide boundaries to the evaluation; separate the relevant from the irrelevant; provide a set of "empty containers" for insights and ideas; provide ideas about samples and sites to be used and about the format and content of instruments.

5. *Sites and sources for data collection*

What sites will be chosen for NE? In what order might they be visited? What individuals will provide data? We do not, of course, mean pre-selection of individuals in the sample, but the types of people to be part of the purposive sample will have to be identified. Where do we begin to maximize the range of information? Where do we go later, for depth of information?

6. *How will the evaluators work?*

Will they work as a team? Will there be more than one team? Will they work separately and then match responses as part of data collation? If the latter, how and how often will it be done? When will the member check -- that is, checking back with data providers for correctness -- take place? At one single time or at multiple times?

What will be done about leaving an audit trail? Will auditors accompany evaluators? How will their work relate to the work of evaluators?

7. *Determining methods, and use of recording equipment*

What methods will be used for data collection? Interviews, observations, analysis of documents and records, unobtrusive measures of various kinds, and other qualitative methods are usable. Will pictures, audio-tapes, film or video cameras be used? What will be done with the materials so collected, and when?

8. *Self-training and training of field workers*

Naturalistic evaluators need to prepare themselves as well, especially if it is their first NE. It may not be inconceivable to undergo a group experience of some sort, which enables the trainees to go through some clarification and heightening of sensibilities. More specific skills in interviewing and observation should also be learned. These should be carefully and patiently taught to all those who will act as collaborators of the naturalistic evaluator.

9. *Other logistics*

These will include typical arrangements about travel, stay in the field, handling emergencies of health and other kinds, etc.

10. *Modes of data interpretation*

NE involves collective data interpretation. Systematic attention should be paid to how these collectivities will be brought together and the interpretation process completed. Who will the evaluators negotiate with, and how?

11. *The audit trail*

In the same way, attention should be paid to audit of the study. The evaluators will have to decide upon the things to do so that audits can be meaningful. One of the things to do might be for all those involved to keep a reflective journal. What other records will be kept so that changes in the themes, samples, and so on can all be recorded for later audit?

12. *Report writing*

Proposals should be made about the kinds of report that will be made. Some may be oral, some may be written. The audiences will differ. The proposal should also discuss such questions as whether graphs, charts, matrices and networks, and descriptive statistics should be included.

(Please see also Chapter 14, "Writing a Proposal for an Evaluation Study in the Rationalistic Mode".)

Things to do or think about

1. What are some information needs of your program that can be fulfilled through NE?

2. Write a proposal for an evaluation study in the naturalistic mode. Ask a colleague to criticize it.

Notes

1. Erickson, Frederick, "Qualitative research in teaching." In Wittrock, M.C. (ed.), *Handbook of research in teaching*. (3rd edition) New York, NY: Macmillan, 1986.

2. Those who are familiar with the work of Egon G. Guba and Yvonna S. Lincoln will recognize my many debts to their work. I have also used and adapted many ideas from "Essential Elements in a Naturalistic Thesis Proposal" that Egon G. Guba wrote for use by doctoral students in the School of Education at Indiana University.

CHAPTER 11

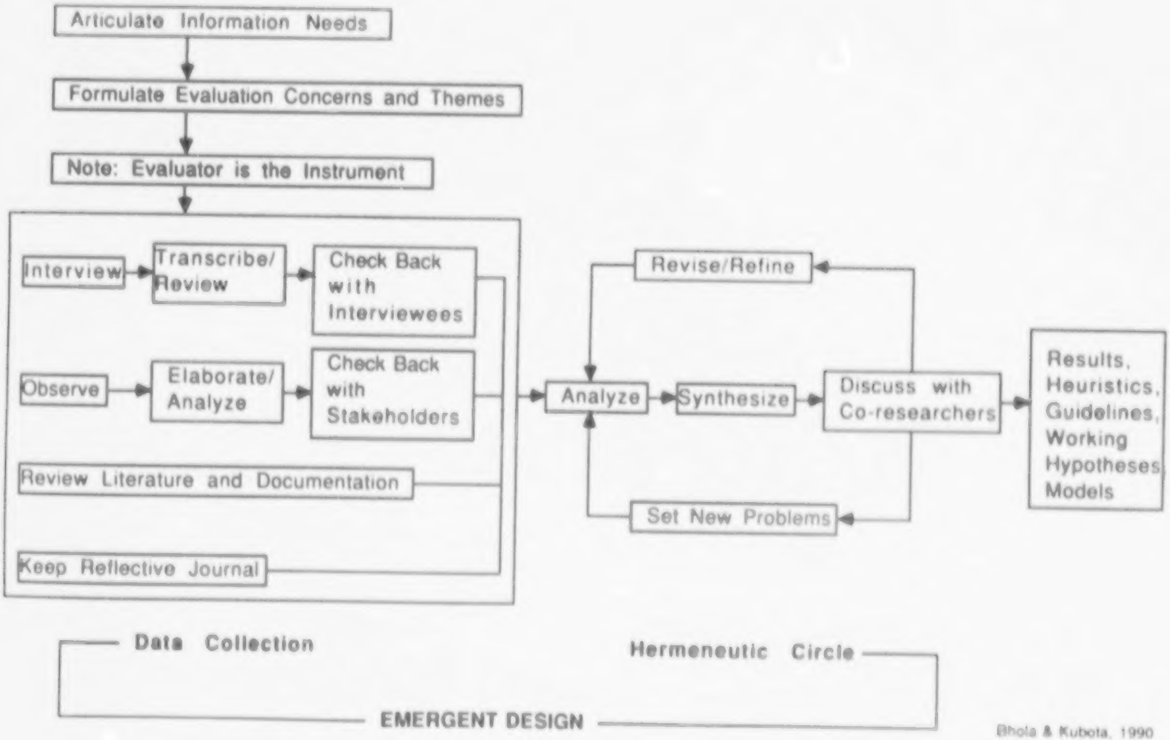
THE PROCESS AT A GLANCE: TOOLS AND TECHNIQUES OF NATURALISTIC EVALUATION

The most important fact of naturalistic evaluation (NE) is that it does not apologize for the use of the human individual as the instrument of data collection. It makes no pretence of constructing "reliable" instruments that collect "objective" data. Indeed, the most widely used instruments in NE are the unstructured interview and participant observation. Documentary materials are also widely used. The naturalistic evaluator does not pretend to stand outside his or her evaluation study. A reflective journal is often used to record personal thoughts and experiences.

Naturalistic evaluation seeks to study life as a whole with all its complexities, as it is rooted in its context, and is experienced by those who are immersed in the reality being studied. It is a personal encounter, not any dispassionate examination.[1] It should not be surprising, therefore, that the naturalistic evaluator wants to see, ask, and interact to piece things together.

The basic instrument in NE is the human instrument -- the evaluator himself or herself. Since NE is interested both in an actor's meanings and the context, there is almost always use of unstructured interview and unstructured observation so that thick descriptions (descriptions that give the reader the feeling of having been there) can be developed. These thick descriptions have to be resonant and coherent. Additional tools and techniques are those of content analysis of documents.[2]

Figure 6 on page 176 graphically presents the typical progression of naturalistic inquiry. As can be seen from the diagram, the data collection phase will typically involve interviews, observation, documentary analysis and a reflective journal. As part of data analysis, this data will be subjected to the sub-design of the hermeneutic circle to come up with required results in terms of the study, and/or heuristics, guidelines, working hypotheses or models.


Articulate Information Needs
Formulate Evaluation Concerns and Themes
Note: Evaluator is the Instrument
Interview
Transcribe/ Review
Check Back with Interviewees
Observe
Elaborate/ Analyze
Check Back with Stakeholders
Review Literature and Documentation
Keep Reflective Journal
Revise/Refine
Analyze
Synthesize
Discuss with Co-researchers
Set New Problems
Results, Heuristics, Guidelines, Working Hypotheses, Models
Data Collection
Hermeneutic Circle
EMERGENT DESIGN

Bhola & Kubota, 1990

Figure 6: The Process of Naturalistic Evaluation (NE)

Methods of data collection

In the following, we include notes on some typical tools of NE.

Interview in NE

The NE interview is unstructured in the sense that it does not ask the interviewee a set of standard, ready-made questions, but it is structured in a deeper sense and at a higher level. Such an interview could be described, in Paulo Freire's words, as dialogic. The purpose is not to get some answers to questions, but to enable the respondent to describe his or her world in his or her own words. The interviewer by whispering genuine questionings, by making thoughtful comments, and by providing reassurances and reinforcements, enables the interviewee to get in touch with his or her inner self and to formulate his or her own meanings of the surrounding realities with coherence.[3]

The interviewer starts by posing general themes and exposing the problematic roots of these themes. After establishing the general boundaries of the subject of the interview, the naturalistic interviewer lets the interviewee take over. The interviewer listens with interest and sympathy, encouraging the interviewee to go on, to explain further, to come back to the point, to choose, to judge, and to take positions. Such interviews typically last several hours. Naturalistic evaluation is indeed a labor-intensive affair.

In-depth interviews may be recorded by the interviewer on paper in the presence of the interviewee or immediately after the interview when the interview is still fresh in the interviewer's mind. The tape-recording of the interview may be preferable, if the interviewee willingly permits this. (For a discussion of *structured* interviews, see the section on "Tools and Techniques of Rationalistic Evaluation" in Chapter 15.)

Focus group interviews

The naturalistic evaluator can often make good use of group interviews focused on particular issues. The focus group interview technique has been in use in marketing research in America for over twenty years. The objective of focus group interviews is to gather in-depth information through group discussion, thereby getting at the thoughts, perceptions, feelings and attitudes of persons knowledgeable about a program.

A focus group is typically a homogeneous group of seven or more people in "directed" interaction with each other. The direction of the group is handled by a moderator. An observer is also present. Tape recordings of group interactions are typically made. Discussion begins with a written list of concerns, also called "stimulus questions". Focus group interviews take 45 to 60 minutes to conduct.

As in all group dynamic techniques, the moderator has to direct the group in such a way that enough ideas are produced; that a variety of ideas are produced; that ideas produced are of high quality; that participants do not get side-tracked; that spontaneity is maintained and ideas are shared even when they are not well developed; that people produce ideas and do not simply react to ideas produced by others; and that the personality and status of participants do not come into play in the group.[4]

Field observation

Field observation is a data collection strategy that can be used both for rationalistic and naturalistic evaluations.[5] In the NE tradition observation will be unstructured and leisurely. It can be either participant observation or nonparticipant observation.

In observation, the phases of recording and interpreting should be separated. The observer should record what he or she saw. What it could have meant should be saved for later interpretation. Whether it is used in NE or RE, systematic instruments can be developed in each case for recording observation data. (See further discussion of Observation under "Tools and Techniques of Rationalistic Evaluation" in Chapter 15.)

Doing tracer studies and making chronologs

A tracer study, as the name suggests, traces the path of progress of a person or persons over time. Thus, a tracer study of a new literate will read like a short biographical sketch of that person, recording what literacy may have done to the new literate's life as he or she has put literacy to various uses in life and work.

A colleague of the author at Indiana University, Professor Myrtle Scott, has developed the concept of *chronolog*, which involves following subjects around for a particular period of time -- an hour, a day, a week -- in their natural habitat (a village community, a school, a hospital) and recording what they do. Such chronologs would be particularly useful for the study of the emergent roles and functions of development agents and literacy workers at the field level.

Analysis of records and documents

Analysis of records and documents can itself be conducted in the RE or the NE mode.[6] In the RE mode, content analysis involves random selection of content and statistical techniques for making general assertions. In the NE mode content analysis is more concerned with the aggregation of meanings and the crystallization of themes which may be embedded in various documents.

Unobtrusive measures

Unobtrusive measures,[7] again, can be used both for RE and NE studies. As their name suggests, unobtrusive measures do not impose on the respondents. One watches the behavioral footprints they leave behind. The condition of the literacy primer in the hands of the learner may tell how much of it has been read. The way the field worker dresses may indicate his or her "social distance" from the people he or she seeks to serve. The garbage generated by a family may tell us a lot about their shopping and consumption habits -- unobtrusively!

The reflective journal

The reflective journal is exactly what the name suggests -- a journal in which reflections are recorded. (Lincoln and Guba use the term "reflexive journal" to point out that the evaluator is recording in the journal a lot of information about *self* as a human instrument of evaluation.) The evaluator keeps a journal (a diary) in which he or she records, preferably on a daily basis, (i) the daily schedule and the logistics involved in the study; (ii) notes on the day-to-day methodological decisions and why such decisions were necessary; and (iii) his or her personal reflections on experiences and anticipations, clashes of values and collaborations with stakeholders, the boredom of work and the excitement of emerging insights. The reflective (or reflexive) journal thus should become the super-ego of the evaluator, an aid to memory and a source to return to for checking interpretations.

Cases of NE methods in action

The theory and methodology of NE is still in the process of discussion and emergence. A variety of methods, such as those discussed above, are typically combined as long as the underlying assumptions are those of naturalistic inquiry (that is, reality is a social construction) and the objective is not to make law-like universal statements, but "warranted assertions" in particular contexts.

There are few naturalistic evaluations available in the field of literacy that could be presented as models to follow. It would be useful nonetheless to outline the methodological approaches and methods used by some educational evaluators who claim to conduct naturalistic analysis or to have a naturalistic orientation. Two cases are therefore outlined below on pages 182-184.

Readers should be able to develop from these some initial ideas as to what it means to do an evaluation study with the NE orientation.

NE: Data analysis

Data analysis in NE is, in some ways, a much more challenging process than statistical data analysis. One is involved not merely in the aggregation of numbers, but in the generation of meanings, and in the search for larger patterns in which such meanings reside.

The first step in NE data analysis is total immersion in the data already collected. The evaluator must read and re-read the transcripts of interviews, reports on observations, notes on documentary analysis and the reflective journal. (See Figure 7 on the next page.) Key words and phrases, and recurrent themes, should be written on cards, as also the significant quotes from remarks made by various stakeholders. Through a process of synthesis, using the CLER model and the "Before and After" format, changes in the lives of individuals, groups, institutions and communities should be reported. There are, of course, no standard formulas in NE data analysis, but evaluators using this mode are sure to gain from experience as they try to make sense of the world on the basis of the data collected.

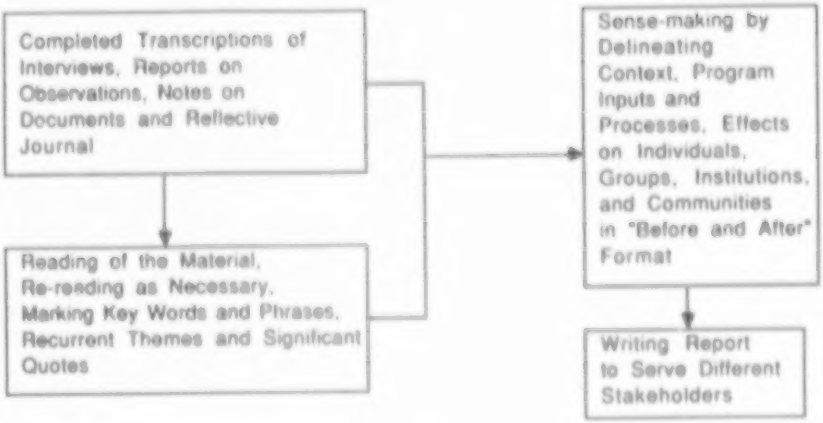

Completed Transcriptions of Interviews, Reports on Observations, Notes on Documents and Reflective Journal
Reading of the Material, Re-reading as Necessary, Marking Key Words and Phrases, Recurrent Themes and Significant Quotes
Sense-making by Delineating Context, Program Inputs and Processes, Effects on Individuals, Groups, Institutions, and Communities in "Before and After" Format
Writing Report to Serve Different Stakeholders

Figure 7: NE - Focus on Data Processing and Analysis

CASE 1

John W. Creswell and his six associates[6] did a naturalistic analysis of the faculty development role of department chairs (chairmen or chairwomen) in higher education settings. These inquirers cannot be considered "purists" in NE methodology but did view their work as "a process of research with strong naturalistic orientation, an orientation of inductively isolating a problem for study, discovering rather than verifying an *a priori* theoretical framework, and descriptively reporting results to date, subject to later qualitative exploration and quantitative investigation".

The research method of naturalistic analysis as operationalized in this study consisted of four phases:

Phase 1	Establishing direction for the research
Phase 2	Characterizing the sample and developing interview procedures for the national study
Phase 3	Conducting the telephone interviews
Phase 4	Analyzing the verbal report data obtained from interview cases

Detailed actions in each of the above phases were as follows:

Phase 1

Mapped the dimensions for studying the chair role. This included some general questions such as: What are the demographic characteristics of "effective" chairs and do they feel responsible for assisting faculty growth and development? What kinds of faculty situation call for their assistance, and what approaches do chairs use? How is chair assistance shaped by contextual variables in the academic workplace, such as career stage of the faculty member; whether assistance is initiated by the chair or faculty member; discipline differences and institutional differences?

Used 99 activities that chairs had been found to engage in.

Developed a concept paper.

Nationally recognized leaders in faculty development reacted to the concept paper.

Conducted a pilot study which also gave opportunities for testing telephone interview techniques.

Two doctoral dissertations were directed.

Together, the above actions resulted in the following: Content areas for questioning began to emerge, such as, background characteristics of chairs; faculty issues/situations; faculty development practices in general; conditions of departments.

Phase 2

Sought representativeness in sample.

Developed a semi-structured interview schedule; and trained interviewers for consistent and accurate administration of the schedule.

Used code forms to record interview responses.

The first cycle interview asked the people to "identify 3-5 department chairs who excelled in assisting faculty in their growth and development".

Chose a random sample of nominators/nominations.

Phase 3

Over a nine-month period interviews were conducted and recorded. Average interview 45 minutes.

Checked for inter-interviewer reliability.

Phase 4

Analysis began as early as sufficient cases were available. The following happened:

First fifty interviews were used to develop analysis procedures.

Open-ended questions from these fifty cases were "forced" into preliminary categories and labelled according to question focus.

Then the remaining 135 cases were incorporated into the analysis.

Various "stories" helped identify types of "faculty situations" and "chair approaches".

Triangulation among research team to recategorize situations/approaches by examining "outlier" cases, maximizing similarities and differences, and by developing prototypes.

Result: a conceptual matrix.

All data used was self-report data.

CASE II

In a recent evaluation of the programs of the Adult Literacy Organization of Zimbabwe (ALOZ) conducted by Bhola and Muyoba,[9] the methodology, once again, had a naturalistic orientation. The process and steps can be delineated as follows:

The evaluation contract had listed general questions that needed to be answered: How had ALOZ adult literacy and income-generating programs affected the lives of those it sought to serve? In what ways did ALOZ literacy work contribute to the literacy promotion efforts at the national level in Zimbabwe? How did the assistance provided by USAID/ZIMBABWE to ALOZ during the early 1980's contribute to the fulfillment of the ALOZ mission?

A detailed analysis was conducted of documentation related to the development strategy in Zimbabwe and the role assigned to literacy in the strategy; the national policy of literacy and the performance of the government's national literacy campaign; the mission of ALOZ in relation to literacy promotion and the materialization of that mission; and the goals and objectives of USAID/ZIMBABWE grants to ALOZ.

A set of 12 tables was designed to develop a numerical portrait of the work and achievements of ALOZ during the period of the USAID grants.

To develop a picture of how the providers of ALOZ programs and the learners and other beneficiaries of the programs experienced the ALOZ program, a considerable amount of field work was undertaken.

A purposive sample of typical localities to cover different language areas (Shona and Sindebele), and different socio-economic realities (urban-industrial areas, manufacturing plants, agro-industries, rural estates, rural settlement schemes, etc.) were covered.

In each locality, the whole range of stakeholders were interviewed, among them learners, their spouses, other family members, indirect beneficiaries of income-generating activities in the villages and communities, teachers, supervisors, trainers, agents of sponsoring and collaborating agencies, etc. Semi-structured interview schedules were used as bases for conversations with respondents.

First, self-contained case studies -- one for each of the localities covered, and using responses from relevant clusters of stake-holders -- were written. Then all data -- numerical data organized in tables, and case studies -- were used in various ways to answer the specific questions asked by decision-makers.

Things to do or think about

1. Use a tape recorder to record your interview with a farmer to find out why he is unable to follow all the advice he gets from the extension worker. Play the interview back to yourself. Who is talking more -- you or the farmer? Are your questions becoming somewhat impatient? Are you really listening?

2. Suppose you are interested in learning about the general level of health in a village community. What will you observe? Compare your observations with the observations of a health worker.

Notes

1. For a general discussion of field research see Johnson, J.M., *Doing field research*. New York, NY: The Free Press, 1975. Discussion of grounded theory and advancements in the concept of grounded theory can be found in Glaser, Barney and Strauss, Anselm L., *The discovery of grounded theory*. Chicago, Il.: Aldine Publishers, 1967; and in Glaser, Barney, *Theoretical sensitivity: Advances in the methodology of grounded theory*. Mill Valley, CA: Sociology Press, 1978.

2. Some references are:
   Bogdan, R. and Biklen, S.K. *Qualitative research for education*. Boston, MA: Allyn and Bacon, 1982.
   Patton, Michael Quinn. *Qualitative evaluation methods*. Beverly Hills, CA: Sage, 1980.
   Miles, Matthew and Hubermann, A. Michael. *Qualitative data analysis: A source book of new methods*. Beverly Hills, CA: Sage, 1984.

3. Gorden, R., *Interviewing*. 3rd ed. Homewood, Il.: Dorsey Press, 1980 and Dexter, L.A., *Elite and specialized interviewing*. Evanston, Il.: Northwestern University Press, 1970, are some useful references on interviewing. One should also refer to the literature on ethnographic interviews. Ethnography uses description as a fundamental component of data collection. The emphasis is on the emic (insider's) perspective to be able to understand the

reality of social systems; and at the same time it provides the context so that data can be properly understood. See Fetterman, David M., *Ethnography in educational evaluation*. Beverly Hills, CA: Sage, 1984; Spindler, George and Louise, *Interpretive ethnography of education at home and abroad*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc., 1987; and Spradley, James P., *The ethnographic interview*. New York, NY: Holt, Rinehart and Winston, 1979.

4. Krueger, Richard A. *Focus group interviewing: Step by step instructions for extension workers*. Minnesota Agricultural Extension Service (320C Vocational Technical Building, 1954 Buford Avenue, St. Paul, Minnesota, 55108), 1985. Also, Qualitative Research Council of the Advertising Research Foundation, *Focus groups: Issues and approaches*. New York, NY: Advertising Research Foundation, Inc., 1985.

5. McCall, G.J. and Simmons, J.L. (eds.) *Issues in participant observation: A text and reader*. Reading, MA: Addison-Wesley, 1969. Also Spradley, J.P., *Participant observation*. New York, NY: Holt, Rinehart and Winston, 1980.

6. Krippendorff, Klaus. *Content analysis*. Beverly Hills, CA: Sage, 1980.

7. See Webb, E.J. et al. *Unobtrusive measures*. Skokie, Il.: Rand McNally, 1966 for a discussion of these measures. For a more recent discussion see Sechrest, Lee (ed.), *Unobtrusive measures today*. New Directions for Methodology of Behavioral Sciences, No. 1, 1979, San Francisco, CA: Jossey-Bass, 1979.

8. Creswell, John W., et al. "The faculty development role of department chairs: A naturalistic analysis." A contributed research paper presented at the Annual Meeting of the Association for the Study of Higher Education, Baltimore, Maryland, November 21-24, 1987.

9. Bhola, H.S. and Muyoba, G.N. *The Role of the Adult Literacy Organization of Zimbabwe (A) in Promoting Universal Literacy -- A Retrospect and a Prospect*. Harare: ALOZ/USAID-ZIMBABWE, 1989. An interesting description of the implementation of the naturalistic approach is found in Valbuena Paz,

Antonio, and González Olivares, Guido, "Case Study of CESAP Programme: 'Mucuchies Peasant Programme'". Hamburg: Unesco Institute for Education (UIE) project PRG 5.14/4.53, Document 10, June 1990. Mimeo.

CHAPTER 12

WRITING REPORTS OF NATURALISTIC EVALUATIONS
AND
WRITING PERIODICAL REPORTS NATURALISTICALLY

The report of a naturalistic evaluation (NE) study is typically a case study. What was promised in the proposal in a future tense is now written in the past tense, with modifications as they occurred. In any program context, a multiplicity of periodical reports are sent to the headquarters from the field and others are written by officers at the headquarters after their supervisory field visits. The techniques of writing NE reports can be extended to writing all periodical reports from the field, "naturalistically".

Evaluation reports are typically made for informational purposes: to inform decision-makers on the state of affairs in a program and to suggest possibilities for improvements. The informational purposes of reports remain primary in NE as well, but the purposes of reporting are expanded to fulfill the following three purposes:

1. Communication
2. Closure, and
3. Commitment

The NE report must *communicate* information to decision-makers. They must get a fix on the state of affairs. It must, however, also give everyone involved a *sense of closure*. Even though we often talk of evaluation being a continuous process, everyone involved must have a feeling that a particular study is now completed and a particular matter is behind them all.

Finally, the report should push all the various audiences into a phase of action and create an opportunity for them to make public their commitments to action.

The naturalistic paradigm demands that the evaluator when reporting on the results of the evaluation study provide the reader with "thick descriptions". The evaluator conducting a study in the NE mode cannot simply process data and offer results and some discussion thereof. The evaluator must tell the whole story, "rich" in detail. Understandably, the report of a naturalistic evaluation study is a case study.

A caveat should be offered, however. High-level decision-makers in bureaucracies are not the only recipients of reports on naturalistic evaluations. Naturalistic evaluations are not only rooted in a new epistemology, but are also responsive to a particular social ethic. The naturalistic evaluator seeks to address all the stakeholders, and particularly the powerless.

Some of these stakeholders (in developing countries as well as in the developed societies) may not always be able to receive evaluation reports in print. This means that some of these evaluation reports will have to be made verbally or as audio-visual presentations.

Making effective presentations of evaluation results

A lot is known about making effective oral speeches. One cannot, however, say too much within the scope of this small handbook. Only a few suggestions can be made.

Oral reporting

Needless to say, the presenter of the report should be well prepared. What are the findings that must be shared with the group? What are the understandings you want the group to develop? What are the likely misunderstandings that must be avoided? The main points of the report should be written down on paper by the presenter and kept in hand.

While the presenter should be well prepared, he or she should not plan for a flawless uninterrupted performance before a tongue-tied audience. People should be allowed to comment, ask questions, raise doubts and ask for discussion of aspects the presenter may not have initially intended to offer. The list of ideas prepared earlier should only be used as a check-list to ensure that all ideas are covered.

The presenter should, of course, speak clearly and audibly and present the report with courtesy and patience.

Making audio-visual presentations

The making of audio-visual presentations has also been reduced to an art. Regrettably, not much can be said on this here, but a few comments will be made.

First, choose each of the different media for the special contribution it can make. Choose a chart when you need to show some important facts and their relationships and you want these to stay in front of the audience for a long time. Choose a film when you want to show motion, and a model when you want them to experience something in three dimensions.

Second, once you have put some media in the presentation, let them work. Too many people will display a chart but not even refer to it in their presentation. Others will project a film and later not integrate it within the rest of their presentation. Once again, the logistics of media utilization should be properly managed. There should be some way to put the chart on the wall, and the model on the table; the film projector should work and the film should not keep on breaking or jumping; and due attention should be paid to darkening and ventilation in projection rooms.

Writing case studies

A good case study tells the whole story systematically, clearly and intelligently. The best advice on writing a case study would be to put in the past tense what was promised in the future tense in the proposal stage of the evaluation study, reflecting and explaining the changes and modifications made.

The case study written to report on a naturalistic evaluation, will use historical-chronological organization. The CLER model discussed earlier should be embedded in the case study. The case study should be full of vignettes and actual quotes from respondents. It does not mean, however, that a case study on NE will include no numbers, tables or matrices. It may or may not, depending upon the area of study and available data.

Writing field reports naturalistically

Within the context of a development program, numerous periodical reports are written: some are sent by the field staff to the headquarters, others are written by officials from the headquarters after their supervisory visits to the field. We strongly suggest that these reports be written in the mode we have described as naturalistic. This will not happen, of course, unless the program officials begin to look at themselves as professionals rather than policemen. At one level this will involve a revolution in the norms of the practice of development. The understanding will have to emerge that government intentions do not determine development. People develop themselves. If people do not become motivated, no development will occur. In addition, there are other circumstances beyond the control of governments and their functionaries. While there are functionaries in the field who are uncommitted and corrupt, and need to be policed and punished, it is not every time their fault, if development does not come about. We need to look at development professionally, as a process which is complex and needs creative responses. Related to the above is the idea that the functionaries at various levels of government have to get out of their obsession with superordination-subordination and develop colleagueship among themselves.

Reports from the perspective of officials at the HQ

Reports written by officers at the headquarters after their field inspections are particularly amenable to being written in the naturalistic mode. Here are some hints:

1. Think of evaluation as a continuous process. Build each new visit on the last one. Study your own earlier reports before embarking on a new trip and make notes about the things to look for.
2. Review the context of the project and your own visit. Think of the themes you will pursue this time in the field. Remember that you will never learn much about reality without encountering reality. That means that you do not just visit offices and look at official files and registers. Identify what you will personally observe -- classes, homes; and what people you will interact with and hear from -- farmers, chiefs, literacy teachers, so-called dropouts.
3. During the visit write a journal with thick descriptions. Check your understandings and perceptions with others and particularly with those to whom they pertain. Suggest

methods for the amelioration of problems and make commitments to do your part.

4. Back at headquarters, after the visit, write a report as suggested above in the case study manner. Share it with your colleagues.

Part V

Evaluation in the Rationalistic Mode

We first introduced the paradigm of Rationalistic Evaluation (RE) in Part I, Chapter 2, of this handbook. Another brief description of RE was included in Part II, Chapter 4, where RE was presented as one of the three components of the methodological triangle of evaluation: MIS, NE and RE.

We have suggested earlier that MIS should be considered to be the most important, indeed the solid base, of the methodological triangle of evaluation. We have suggested further that the second priority should go to NE, which is most suited to discovering qualitative meanings that programs and projects may have had for those whom they have sought to serve. To our present thinking, therefore, RE has the third place in evaluation management.

Third place for RE does not, however, mean no use. We do not by any means suggest that RE has no role to play in the evaluation of educational and developmental programs. In fact the triangulation of the various evaluation methodologies is implicit in the very label of our model: the methodological *triangle* of evaluation. What we are saying is that RE should not be selected simply because, for many many years, it has been mistakenly considered to be the only "scientific" approach to evaluation and research. The MIS and NE take priority.

With the above caution in place, let us suggest that there will be multiple opportunities for evaluators of educational and development programs to practice RE. Let us be reminded of Cronbach's concept of the "context of control" discussed earlier in the handbook. There are indeed many contexts where assumptions of control over the reality being studied can be made and, therefore, the rationalistic paradigm of evaluation can be used without doing violence to the actual reality. In such cases, the use of RE would make considerable sense.

Our discussion of RE will be divided into the following chapters:

13. Rationalistic Evaluation -- Theory, Questions and Design

14. Writing a Proposal for an Evaluation Study in the Rationalistic Mode

15. The Process at a Glance: Tools and Techniques of Rationalistic Evaluation:
    Section A: Tools and Instruments
    Section B: Data Collection
    Section C: Processing and Display of Data
    Section D: Statistical Analysis of Data, and

16. Writing Reports on Rationalistic Evaluations and Promoting Utilization of Results.

CHAPTER 13

RATIONALISTIC EVALUATION -- THEORY, QUESTIONS AND DESIGN

Rationalistic evaluation (RE) makes a particular set of assumptions about reality that include reductionism (that complex social reality can be reduced to simpler aspects for study) and universalism (that universal laws of human behavior can be found that will hold true independently of context). Related to these assumptions is the concept of experimental treatment that enables the researcher or evaluator to fit reality into the evaluator's experimental format, thereby promoting validity, reliability, objectivity and generalizability of results. There are specific evaluation designs and sampling procedures that are part of the theory as well as statistical procedures that make inference from the specific to the general possible with given levels of confidence.

Successes of logical positivism (or the rationalistic paradigm) have been spectacular. Medical researchers, using this paradigm, have banished many deadly diseases and plagues from the face of the Earth; and physicists have put a man on the moon.

Social scientists, to partake of the glory, mimicked the physical scientists and started using the so-called scientific paradigm, almost to the exclusion of anything else.

The magic of the positivist paradigm is finally breaking, and we are beginning to understand that social reality does not fit the rationalistic paradigm very well. Individual behavior does not always tell us much about behavior among groups or within organizations. There are "emergent" properties within wholes which cannot be explained in terms of constituent parts. Conversely, we are understanding that complex phenomena cannot be reduced to simpler aspects for study and then put together as if nothing was lost. The very nature of these phenomena changes as these are fragmented and factored through such reductions.

We are also beginning to appreciate the limits of generalizations. Social phenomena, we now understand, are sensitive to the context in which they take place.

Theory of RE

Proponents of RE have not, of course, surrendered their arms and gone home. While they are beginning to accept the role of judgement in RE and have accepted the existence of problems with validity, reliability, objectivity and generalizability, they still believe that RE according to a "re-conditioned positivism" is the best approach to making normative statements. The definition of reality accepted by RE and the accompanying methodology that is typically proposed has been discussed at some length in Part I, Chapter 2 of this book. A recollection of the essential assumptions of RE is necessary at this point.

RE accepts the existence of objective reality out there for everyone to see. Therefore, RE accepts the possibility of normative statements that are universal and, thereby, generalizable to all settings. RE is built upon the concept of reductionism, which means that the complexity of real life can be reduced to simpler relationships -- individual factors and variables which can be studied in linear relationships to demonstrate correlations or causalities. The assumption is that after being so studied, they can be put back together to help us understand complex relationships. Since causal relationships can be thus established, predictions can be made as well about behavioral events in the future.

The methodology of RE is based on the assumption of control. The evaluator seeks to establish an experimental setting wherein the respondents are selected, treatments are standardized, data collection is objective, and data analysis is typically statistical. These kinds of assumption can be fulfilled under what Lee Cronbach has called the "context of control". Thus there will be questions on literacy and post-literacy, not too many perhaps, to which RE will be the best approach to finding the answers.

Questions in RE -- in the context of control

In our conception, RE seeks to make normative statements about reality that can serve as general guides in a variety of contexts. Taking random samples of individual respondents (or other social units), it seeks to correlate, to compare and to predict at particular levels of confidence.

Normative assertions

(Answers will be based on random samples.)

1. What is the percentage of illiteracy among women in the Southern region of Kenya?
2. What is the rank order among motivations expressed by men for attending literacy classes in a particular program?
3. What is the profile of uses of literacy given by males and females in ages between 30-45 years?

Establishing connections and correlations

(These questions are quite similar to those listed under MIS. The essential difference would be that a practitioner of RE would collect data from a random sample and try to meet the statistical assumptions necessary for making inferences beyond the program population.)

1. What is the correlation between literacy and numeracy skills?
2. What is the correlation between teacher qualification and learner achievement?
3. What is the nature of correlation between literacy score and economic productivity?

Making comparisons between groups and other entities

(Once again these questions look very similar to those listed under the section on MIS. The difference once again is that the practitioner of RE would collect data from random samples and meet other statistical assumptions necessary for making inferences beyond the program population.)

1. What are the differences in achievements in literacy, functionality and awareness between groups of male and female learners from families living on subsistence agriculture and belonging to the same age group?
2. What is the difference in the effectiveness of teachers trained for teaching in the primary schools and new

literates trained within the literacy project to teach adult literacy classes?

3. What is the relative effectiveness of sets of instructional materials prepared according to the Freirean strategy and the whole language approach?
4. What are the distinctions between participants and non-participants in literacy programs on several modernization measures such as economic productivity, nutritional status of the family, family planning, and political participation?
5. What has been the nature and significance of change in the community before and after the implementation of program A?

Design in the RE paradigm

In the dictionary meanings of the term, to design is to develop a conception of something, or is to prepare preliminary plans or sketches for something. In this sense of the word *design*, all evaluation studies must have a design. We must have a conception of what we want to do, why, and we must make some preliminary plans about how to go about doing what we want to do.

In the literature of research and evaluation, however, design has a highly technical meaning. In the RE paradigm, design typically means "experimental design". There has to be a sampling plan, and random samples must be obtained. Evaluation variables must be defined. Evaluation variables must be controlled through various mechanisms. Treatments should be well defined and applied selectively to chosen samples. Instruments are often structured, and statistical techniques are applied to the analysis of collected data.

It is beginning to be understood, however, that "true" experimental designs are seldom possible in education and development. Random samples do not always make sense when dealing with special categories of subject, in particular community contexts. Control of variables and treatments is often impossible. Evaluators are, therefore, now being offered "quasi-experimental designs" -- evaluation designs that are half-way experimental. In using quasi-experimental designs, we try random assignment of treatments, if possible, but control *when* the data will be collected and *from whom*.

Reliability and validity

Researchers and evaluators working within the RE paradigm swear by reliability and validity.

Reliability applies to a test or another measuring instrument. It is defined as a reasonable consistency in results obtained in a sequence or group of repeated tests and measures. A reliable test is one which gives consistent results in different applications to the same subject within a reasonable time-frame. Or, it is one which performs consistently when used by different evaluators, with different subjects. Reliability is necessary though not sufficient for validity.

Validity is the extent to which a test measures the thing it is supposed to measure. Support for validity may be logical or empirical. The test items may have been properly derived from accepted premises by rules of logic; or assumptions may have been based on supportable empirical evidence.

Internal and external validity

The concept of validity not only applies to tests and instruments but also relates to the more general concerns of evaluation design. The results of an evaluation study and the conclusions drawn from these results must be seen as warranted, convincing and acceptable -- that is, they must be seen as valid.

Listed on the following page are some of the assertions that evaluators could make on the basis of their studies, and at the possible objections that could be raised to the validity of such assertions.

ASSERTIONS BY EVALUATORS	*OBJECTIONS TO VALIDITY*
The trainee group has shown considerable learning, as evidenced by the high level of performance on the final test.	Maybe this group was familiar with the content of the training course even before joining it. Maybe the test was easy or the grades have been inflated.
Adult attitudes towards literacy have changed drastically because of the project.	Maybe they have changed not because of the project, but because of the President's speech on national radio. Maybe they have changed not because of the project, but because the newly-opened textile factory has declared its preference for literate and semi-literate labor.
The group of farmers who undertook leadership training at the training institute have assumed actual leadership roles in the community more often than those farmers who did not join leadership training.	Maybe the farmers who undertook leadership training were already in leadership positions and wanted to increase their effectiveness as leaders. Maybe the farmers who joined leadership training were a self-selected group, fired with the ambition to capture the new leadership positions opening up in their communities. Maybe the other group of farmers that is not doing well, is different from the successful leadership group in important socio-economic characteristics, and is thereby disadvantaged.
The farmers' training course increased the overall productivity of farmers who attended by 15% in a year.	Maybe the productivity increase for these farmers last year was 20%. Maybe similar farmer groups elsewhere have shown similar increases.
The introduction of the role of the Family Health Education Worker has changed the level of health in the selected communities from "Poor" to "Medium".	Maybe this is because of the heat and drought of the last year that killed all mosquitoes; and the famine relief high-protein food aid that was provided to families in the area.

These are some examples of the assertions that could be made and the challenges to their validity. Professors Donald T. Campbell and Julian C. Stanley[1] have listed twelve different threats to the internal and external validity of evaluation studies. Evaluators should find their list most instructive:

(A) Internal validity

1. *History.* An outside historical event, such as a presidential speech, or the enthusiasm generated by a newly announced economic plan could challenge the validity of the evaluator's claims.
2. *Maturation.* Individuals being tested as part of the evaluation may mature and grow in such significant ways that they may behave like different people by the time an evaluation study is completed.
3. *Testing.* The first test may teach the items on the test and other related and implied information. The same test (or an equivalent second test) may not then measure real changes brought about by the program.
4. *Instrumentation.* There may have been no changes in the reality but only in the calibration of instruments studying that reality. Or, different observers and examiners may have given different scores for the *same* unchanged reality.
5. *Statistical regression.* This is a statistical phenomenon. Extremely high or extremely low scores on a first test tend to move towards the mean of total scores during a second test. Thus, changes in the scores on a second test may really have nothing to do with respondent groups, program methods, or program effects. Statistical regression occurs specially in cases where groups have been selected on the basis of extreme scores.
6. *Selection.* Biases in the selection of learners for training, interviewing and testing may threaten the validity of results.
7. *Experimental mortality.* Those initially covered by an evaluation study may cease to be participants in the evaluation. They may drop out of the program or may move away in search of food or work. Thus, the residual group may no longer be representative of the group or community being studied.

8. *Selection-maturation interaction.* The peculiar chemistry of the selection process of subjects in an evaluation study and their maturation together may show effects independently of the program inputs and processes.

(B) External validity

9. *The reactive and interactive effect of testing.* The pre-test may increase or decrease the sensitivity or responsiveness of the respondent to certain program treatments applied as part of the evaluation.
10. *Selection-treatment interactions.* The peculiar chemistry of selection of respondents and the instructional and organizational treatments may create effects that falsify results regarding real program effects.
11. *Reactive effects of experimental arrangements.* Persons and groups show one set of effects of a treatment within the experimental setting, but not in non-experimental, real-life settings. Or, in some cases, experimental conditions may be much too artificial.
12. *Multiple-treatment interference.* When the same group is frequently tested, or interviewed many times in different connections, results may become confused. Effects of a test and an interview cannot be erased from the minds of respondents, and the first test or interview may influence later testing and interviewing in ways that we do not understand.

The purpose of evaluation design is to reduce the above mentioned threats to the validity of evaluation results.

Some ideas on sampling

The validity and general rigorousness of evaluation studies can be increased by following proper sampling and design methods. We begin by presenting some simple ideas on sampling.

A *sample* is a portion, part or piece taken or shown as a representative of the whole. Sampling is often a practical need. Evaluators may deal with programs with broad scope, covering hundreds of thousands of people. They cannot go to each and every member of their populations and ask them the questions to which they want answers. Instead they want to select a small number of respondents in such a manner that the sample is representative and can be studied to make inferences about the whole.

We should explain the two words population and representativeness used in the paragraph above. In the everyday meaning of the term, population covers all the people -- men, women and children, young and old, farmers, workers and housewives -- living in a particular community or nation. For the evaluator, *population* is the total group of people in which the evaluator is interested. It may be all women of child-bearing age in a country, all people suffering from lung diseases, all textile workers or all new literates in a region or a township. Samples are drawn from such populations.

Samples have to be *representative,* that is, as parts they have to represent the whole from which they are drawn.

There have been many advances in sampling theory. Statisticians have worked out formulas whereby they can test the representativeness of their samples and calculate the probabilities of error.

Size is an important consideration in selecting samples. Clearly the perfectly representative sample of a population is the population itself. Generally speaking, the larger the sample, the more representative it will be of the population. But unnecessarily large samples will not be good samples. We have to have the right size of sample that is both economical and representative.

On page 205 we have reproduced a table that can be used for determining sample sizes for various population sizes. Let us also look at some frequently used types of sample.

Random sampling

A random sample results when selections are made purely on the basis of chance, without any underlying system or pattern, and when each item or person in the population being studied has had an equal chance of being included in the sample. Random samples of appropriate size are most likely to represent all the characteristics and exact distribution of the total population to the evaluator. One method of taking random samples is to arrange the population in some way, assign numbers to it, and then draw some numbers randomly. Where the populations are big and the numbers to draw from are large, printed tables of random numbers can be used.

Random sampling may often be applied sequentially in evaluation studies. Geographical regions of a country may be selected randomly, followed sequentially first by the random selection of communities within the randomly selected regions, and then by the random selection of adults in the randomly selected communities. Again, randomly selected adults could be assigned to different learner groups through subsequent random selection.

List sampling

List sampling is a modification of the random selection method. The population of interest to the evaluator is arranged in a list according to some rule -- alphabetically, for example -- and then every nth number is selected from the list. For example, every 5th or every 20th number may be picked, depending upon the size of the population and the size of the sample being selected. The starting point in the selection process can itself be randomly selected to meet the criterion of equal chance of selection for each unit.

Area sampling

In area sampling, some geographical locations may be randomly selected from all available sites, and then all appropriate units within the selected areas may be studied.

Stratified sampling

The population of interest to an evaluator may be divided into distinct socio-economic strata. Or, the population may be stratified according to age groups -- children, young, middle-aged and very old. In such cases, stratified sampling may be used. In accordance with proportions in the total population, samples may be drawn proportionately and randomly from each of the population strata.

Purposive, theoretical or elite sampling

The naturalistic evaluator or researcher may often need not a random sample but a purposive sample, a sample that fulfills his or her particular pre-determined needs. The evaluator may be interested not in any randomly selected group of adults in a community, but in two or three people who are supposed to serve as the community's gate-keepers. The evaluator may be interested, that is, in small elite samples.

TABLE FOR DETERMINING SAMPLE SIZE FROM A GIVEN POPULATION[2]

N	S	N	S	N	S
10	10	220	140	1,200	291
15	14	230	144	1,300	297
20	19	240	148	1,400	302
25	24	250	152	1,500	306
30	28	260	155	1,600	310
35	32	270	159	1,700	313
40	36	280	162	1,800	317
45	40	290	165	1,900	320
50	44	300	169	2,000	322
55	48	320	175	2,200	327
60	52	340	181	2,400	331
65	56	360	186	2,600	335
70	59	380	191	2,800	338
75	63	400	196	3,000	341
80	66	420	201	3,500	346
85	70	440	205	4,000	351
90	73	460	210	4,500	354
95	76	480	214	5,000	357
100	80	500	217	6,000	361
110	86	550	226	7,000	364
120	92	600	234	8,000	367
130	97	650	242	9,000	368
140	103	700	248	10,000	370
150	108	750	254	15,000	375
160	113	800	260	20,000	377
170	118	850	265	30,000	379
180	123	900	269	40,000	380
190	127	950	274	50,000	381
200	132	1,000	278	75,000	382
210	136	1,100	285	100,000	384

Note: N is population size; S is sample size.

Some simple designs for evaluators

Some designs of interest to evaluators working in the RE mode are presented below. A few of these designs may be usable in NE as well. These descriptions are based on the work of Campbell and Stanley referred to earlier.

(i) *The one-shot case study*

Campbell and Stanley call it a pre-experimental design. There is a total absence of control. A program treatment (X) is followed by observation (O):

X O

While a case study implicitly compares its results with similar events casually observed or read and remembered, the case study can be strengthened by more systematic comparisons. At least one more comparison should be attempted. We should remember that this so-called pre-experimental design can be a useful tool of the naturalistic evaluator.

(ii) *The one-group pretest-posttest design*

This is also considered a pre-experimental design and can be represented as follows:

O1 X O2

A first observation or pretest O1 is followed by program treatment (X), after which a second observation or post-test O2 is recorded.

Evaluators in the RE mode will often be using this design in their evaluation studies. They should, however, do their best in defending their results against threats to their validity; or in qualifying their conclusions in the light of effects of history, maturation, testing or instrumentation as discussed above. (We have earlier discussed twelve threats to the internal and external validity of evaluation results. It will be a good idea for evaluators to develop the habit of checking their results in regard to each of these twelve threats, every time they design or complete an evaluation study.)

(iii) *The static-group comparison*

This is a design in which a group which has been subjected to a program treatment is compared to another that has not been:

X O1 O2

This is also a design under many threats of validity. The most obvious ones are those of selection (the two groups may have been different to begin with), and mortality (subjects in the experimental group or the comparative group may have left the groups for some reason).

(iv) *The pretest-posttest control group design*

Campbell and Stanley call it a "true" experimental design. Two samples (R1 and R3) are randomly selected from the same population. One is assigned a program treatment and the other is not:

RO1 X O2
RO3 O4

This design meets most of the standards of internal validity quite adequately, though care must be taken in generalization of results to the general population.

(v) *The posttest-only control group design*

This is another example of the true experimental design. The pretest suggested in the design immediately preceding may not always be possible. It is *not* even necessary, if randomization in group selection can be assured. The design then takes the form:

R X O(1)
R O(2)

(vi) *Quasi-experimental designs: The time-series experiments*

The time series design involves periodic measurement of some individual or group both before and after the introduction of some program treatment and the study of the "discontinuity" introduced in the pattern of behavior in time:

O1 O2 O3 O4 O5 O6 O7 O8

The evaluator using this design must specify in advance the expected time relationships between the introduction of a program treatment and the manifestation of its impact. The relative isolation of the group from outside influence should be ensured as well as some consistency in the conditions.

The above design can be strengthened by working with two groups in a time series as follows:

$$\begin{array}{ccccccccc} O & O & O & O & X & O & O & O & O \\ \hdashline O & O & O & O & & O & O & O & O \end{array}$$

(vii) *Quasi-experimental designs: The nonequivalent control group design*

This is a design in widespread use because it fits the realities of the world of education and development which are often faced. Too often evaluators have to work with already formed groups and classes and cannot assign members to them randomly.

Thus, the design takes the form:

$$\begin{array}{ccc} O & X & O \\ \hdashline O & & O \end{array}$$

We should note the similarities between this quasi-experimental design and the "pretest-posttest control group design" which was described above as a true experimental design. The essential difference between the two designs is that in the case of the "pretest-posttest control group design" the treatment and the control group are chosen randomly while in the "nonequivalent control group design" discussed here, the groups are not randomly chosen and hence are nonequivalent.

Things to do or think about

1. Examine the conclusions of any evaluation study recently done by a colleague in your training institute or in some other development setting. What are some possible rival hypotheses or explanations for the assertions made by the evaluators?

2. Look at the table of "Assertions by evaluators -- Objections to validity" included in the beginning of this chapter. What kinds of design could have been used in each case to defend the validity of conclusions arrived at by educators?

Notes

1. Campbell, Donald T. and Stanley, Julian C. *Experimental and quasi-experimental designs for research.* Chicago, Il.: Rand McNally, 1963.

2. Krejcie, R.V. and Morgan, D. "Determining sample size for research activities". *Educational and Psychological Measurement*, 30: 607-610, 1970.

CHAPTER 14

WRITING A PROPOSAL FOR AN EVALUATION STUDY IN THE RATIONALISTIC MODE

Rationalistic evaluation (RE) proposals are not only comprehensively elaborated, but are meant to be strictly followed. Hypotheses or questions must be carefully stated. Variables must be properly defined. Treatments must be fully articulated. An experimental or quasi-experimental design should be appropriately chosen. Samples should be properly developed and must be protected from history and attrition for valid results. Instruments must be well designed, and pre-tested. Statistical procedures to be followed should also be decided upon beforehand.

Successful, cost-effective and timely completion of an evaluation study requires considerable forethought and pre-planning. This thinking and pre-planning can be best done within the framework of developing a "formal" proposal for the evaluation study. The process of developing a proposal for the evaluation study can be used to systematize the evaluator's own thinking; to clarify technical, secretarial and material needs of the study; to take stock of available resources; to request and receive consultant help, if necessary, on various aspects of the evaluation study; and to use the proposal as a tool of communication with administrators and interested parties.

As has been mentioned before, evaluation studies in the RE mode will typically seek (i) to make normative statements about populations based on randomly selected samples; (ii) to make comparisons between two groups, or before and after comparisons in regard to characteristics of the same group; and (iii) to establish correlations between characteristics of individuals or groups of individuals.

In the context of an RE, evaluators may collect fresh data, or they may use data already included in the MIS. Indeed, given a good MIS, collection of new data may not be necessary every time an RE study is undertaken.

Taking the example of an evaluation study involving a training program for literacy workers, we shall list the various steps involved in developing an evaluation proposal. A beginner, writing his or her first proposal for an evaluation study, may find it useful to go through the following steps, more or less in the order given. The more experienced proposal writer may be able to jump back and forth to various steps: from step 4 to step 7, to step 10, to step 12 and so on. Again, in the settings of training workshops and seminars of short durations, it may be necessary to focus on some steps and not on others.

It should also be kept in mind that until the final proposal is ready, the various parts of the proposal will require constant review and revision. The development of tools and instruments may require a look back at the indicators chosen for the study. A review of the indicators may require rewriting of the evaluation question and of the justification of the study. Even after the proposal is all done, the realities of the field may demand changes and revisions, once again. One should be mentally ready for these never-ending reviews.

We shall now elaborate and expand upon the various steps involved in writing an RE proposal for the evaluation of a development training program:

1. *The developmental context*

The role and functions of the training institute or the training program to be evaluated should be put within the development context. The training program's contribution to the national effort in the training of manpower for development should be briefly indicated.

If the institution offers a variety of training programs, each different program should be listed, with general objectives of each program indicated separately. In some cases, it may be useful to include the organizational chart of the training institution or program.

2. *The description of the training program in design terms*

First, the general characteristics of the training approach should be recollected, e.g.:

(a) Is the training supposed to be general or specialized?
(b) Does it emphasize process or teaching of knowledge and skills?
(c) Is the training planned participatively or is it pre-packaged?
(d) Is the training offered academic or operational?
(e) Does the training seek to teach entrepreneurial values or communal and cooperative values?

There may be some *other* important questions that could be asked, but the above list should provide a good starting point.

These general questions about training design must be followed by a description of the training program to be evaluated in *system terms* (what we have also called design terms). The four system parameters (inputs, processes, contexts and outputs) should be used to describe the training system in concrete terms and values.

3. *The problem set*

Evaluation problems arise from a lack of information or a lack of understanding. We may have no information or we may have insufficient information on inputs and about the context of our work. We may have less than adequate understanding of the processes and their application within our particular setting. We may have no measure of the quantity or quality of our outputs. These shortcomings together will create a whole "set of problems" in any training program. Indeed, a training institution or a training program is unlikely ever to be short of evaluation problems.

In developing a proposal for an evaluation study, an evaluator should review the whole set of interrelated problems found to be bothersome to program administrators and decision-makers. The evaluator must, however, distinguish between evaluation problems and purely administrative problems. Evaluation problems arise from lack of information and understanding, whereas administrative problems arise from incompetence or deliberate neglect of duty. Administrative problems cannot be solved by evaluation.

4. *The evaluation problem chosen for study*

The evaluation problem chosen for study will have to be one out of the set of problems described under the preceding section, "The problem set". A good problem statement is one that is as concrete and specific as possible:

- In place of the total training effort of an institution or program, it may be preferable to evaluate a specific part of the training effort.
- In place of all aspects of a training effort, it may be preferable to evaluate only some aspects of a training effort.

- It may be preferable to cover a sample of a population rather than the total universe.
- It may be preferable to study the implementation of a training program during a specified time period rather than over the total life of the program.
- It may be preferable to look for specific and concrete effects of a training effort rather than its broad and generalized impact.

We are not suggesting that it is impossible or undesirable to study the broad impact of large-scale training programs in terms of their general and long-term influences on large groups of trainees. All we are suggesting is that, in most RE situations, it is more useful to be specific rather than general.

Whether the evaluation problem is defined in general or specific terms, *ambiguity* is not permissible in RE under any circumstances. The evaluator, in stating his or her evaluation problem, should be most careful with words. The words should mean exactly what is in the mind of the evaluator, nothing more and nothing less, leaving no scope for alternative interpretations.

5. *Justifying the choice of the evaluation problem*

The choice of one evaluation problem from a total "set of problems" cannot be arbitrary. The evaluator should be able to justify his or her choice of the particular evaluation problem. The justifications may range from the political, the programmatic, to the merely possible. An evaluation problem may be justified because the donors want it studied or because the planning department or the president's office has asked for the information. At other times, the evaluation problem chosen may have important policy implications or may produce crucial feedback absolutely necessary for the future planning of a program. Or an evaluation problem may be justified in terms of feasibility -- something that can be accomplished with the minimum of resources even though there might be other more important evaluation questions which should have been tackled first if resources had been available.

6. *Review of available research and experience*

Available theory and research may help an evaluator to define and to clarify the evaluation problem and help in asking the right questions or framing the right hypotheses. Other evaluators, in other training settings, may have asked similar questions. Some experience may be available among administrators and trainers who have worked long in similar training situations. An attempt should be made to collect available knowledge, experience and opinion as part of developing the evaluation proposal. We should learn from other people's experience and should not waste our lives in reinventing the wheel!

7. *Asking questions and sub-questions*

It is important to translate the evaluation problem into a set of questions to be answered or hypotheses to be tested. As we have indicated, in RE, questions and hypotheses will arise from the need to make normative statements, comparisons and correlations. Questions, of course, can be stated as hypotheses and vice versa. One need not, however, state one's evaluation interests both as questions and hypotheses, at the same time. That will be a useless redundancy. Indeed, while doing RE it might be best to work with questions and sub-questions and leave hypotheses alone.

8. *Evaluation models and approaches to be used*

Evaluators working in the RE mode will, of course, choose the *classical* (also called the "scientific") paradigm. Even within this paradigm, however, it may be possible to use different kinds of evaluation model, and different information-gathering approaches and techniques. We should remember that RE can and does sometimes use unstructured instruments to collect qualitative data. However, the data so collected are converted into nominal or ordinal categories and processed and analyzed using positivist assumptions. The methodological choices should be made clear, and related assumptions should be articulated as far as possible.

9. *Evaluation design or steps and procedures*

To have an evaluation design means to do all that is necessary to defend the conclusions of your study from attacks on validity and reliability. In RE, there are standard evaluation designs, each requiring standard sets of procedures for their implementation. The essential problem here then is to choose the right design, and to be familiar with the associated statistical procedures of analysis. Major steps in the conduct of the evaluation study and the procedures to be followed at each step should be outlined in this section of the proposal.

10. *Instruments and tools of data collection*

The proposal for an evaluation study should include a discussion of the tools and instruments that will be used for the collection of data. Preferably the first drafts of the tools and instruments should be attached to the proposal.

There are two prior questions that the evaluator must face before getting on with the construction of the tools and instruments: (1) What is the unit of *analysis*? In other words, where are effects and consequences likely to appear -- in individuals, in families or groups, in organizations, or communities? (2) What will be the *indicators* of effects and consequences having actually appeared? In other words, what responses and behaviors, for example, will indicate change in motivations or in the learning of self-reliance?

The units of analysis should be carefully chosen and proposals should also include suggestions about pre-testing of tools and instruments in pilot settings. Rehearsals are as important for the act of data collection as they are in the staging of a play.

11. *Field work and related research plans*

A proposal for an evaluation study should include plans for library research as well as data collection from the field. If documents or reports will be needed, the evaluator should know where to find them, who will have them, how to obtain copies of those documents, and how much time it might take to obtain them.

Plans for collection of field data should be made carefully. If the evaluator cannot collect all the data personally, investigators or interviewers may have to be hired. This means that plans must be made for their recruitment and training. Local contacts in the field must be identified and orientation must be provided to them about the objectives of research and about research plans.

Field visits must fit the realities of the field and the convenience of individual respondents. The evaluator must keep in mind such considerations as the harvesting season, the weather, fairs and festivals and visits of V.I.P.'s, examination schedules in schools and training institutions, and planning and budgeting cycles in departments and ministries. Problems of transportation should be anticipated and solved. Keeping all of the preceding in view, a time schedule should be prepared.

12. *Plans for data processing and data analysis*

Plans for data processing and data analysis must also form part of the proposal for the evaluation study. Will coding sheets or tabulations be needed for data collation? If so, these should be prepared and tested. Personnel needed for coding and collating data should be recruited and trained. The need for technical consultancy or statistical help (even computer time, if required) should be anticipated and plans made for receiving such help.

As in the case of planning for data collection, plans for data processing and data analysis must also be prepared in terms of a time schedule. Mere lists of things to be done is not enough; plans must be time-sensitive.

13. *Budgetary plans*

The conduct of an evaluation study will need staff time; secretarial and duplication help; paper, postage, tape and tape recorders (in some cases); field investigators; and transportation and telephone costs, etc. All these resources exist within training institutions and programs and should be available to those who want to use them. It is impossible to think of a training institution that would not want its trainers to do the best training job possible. Good training requires feedback; and, therefore, evaluation has to be an integral part of all good training. The resources available in the institution for "training" should be equally available for the "evaluation of training". Trainers-evaluators should use these already available resources rather than always asking for new resources within the context of their evaluation studies. Where new resources are absolutely necessary, a careful budget should be made.

14. *Report writing*

The proposal for an evaluation study should also include the element of "reporting plans". Will the evaluation results be used within the program or the institution, or will they be disseminated outside the institution? If dissemination outside the institution is envisaged, a clear description of outside clients and consumers of the evaluation study should be developed. The same report is not necessarily appropriate for all groups; and writing different versions of the report should be considered.

In writing an evaluation report, the policy and program implications of data should be brought out. Data do not always speak for themselves. While it is necessary that evaluators bring out the implications of their findings for policy-makers and program planners, they should not draw unwarranted conclusions. Opinions and hunches may be offered but should not be mixed with inferences from the data.

Evaluative information can be both used and abused. Too often readers of evaluation studies may be in search of culprits rather than causes; and may want to punish rather than plan with greater understanding in the future. No wonder that colleagues whose work is being evaluated will often get worried about the evaluation process and what it might find. To handle the departmental politics of evaluation, it may be useful to discuss the preliminary report of evaluation in a group setting before issuing a final evaluation report.

Not all evaluation studies need be duplicated and distributed. A single copy of an evaluation study will be worth a thousand, if its findings illuminate action and if its recommendations become part of decision-making.

15. *Bibliography*

A proposal for an evaluation study should also include a bibliography of books, reports and documents used in developing the proposal and likely to be used in the conduct of the study and in writing the final report.

Things to do or think about

1. Prepare a formal proposal for an evaluation study in the RE mode, using an evaluation question of your choice.

2. Have you conducted an RE study before? Do you think your evaluation study could have been improved if a formal proposal had been written before the actual implementation of the study? If you have never conducted an evaluation study yourself, discuss the usefulness of the ideas included in this chapter with someone who has.

CHAPTER 15

THE PROCESS AT A GLANCE: TOOLS AND TECHNIQUES OF RATIONALISTIC EVALUATION

The tools and techniques of rationalistic evaluation (RE) are, understandably, highly rationalized. RE instruments are, typically, pre-structured and pre-tested. Detailed codes are developed for any open-ended questions included in the instruments. The field investigators are advised to be impersonal in order to be objective. Statistical methods and levels of confidence to be placed on the inferences to be made are agreed upon beforehand.

The problems of data collection in the real world are by no means minimized by the rationalization of instruments. Human problems remain that require special attention and which are solved by rationalistic evaluators in their special ways. Once data have been collected, they have to be collated, processed and displayed in special formats for statistical analysis to test hypotheses and answer evaluation questions.

First, it would be useful to be reminded at this stage that the process of "evaluation planning" described in Part II, Chapter 3 above, applies to all the three information-generation approaches: MIS, NE and RE. Second, there are clear parallels between the design of an MIS and the design of RE. The steps peculiar to the progression of evaluation in the RE mode are shown graphically in Figure 8 on the next page.

In this chapter, we shall be discussing the important questions of designing instruments; administration of these instruments in the field; collation (putting together) of data collected from the field; processing and display of data; and finally data analysis to test hypotheses or to answer questions.

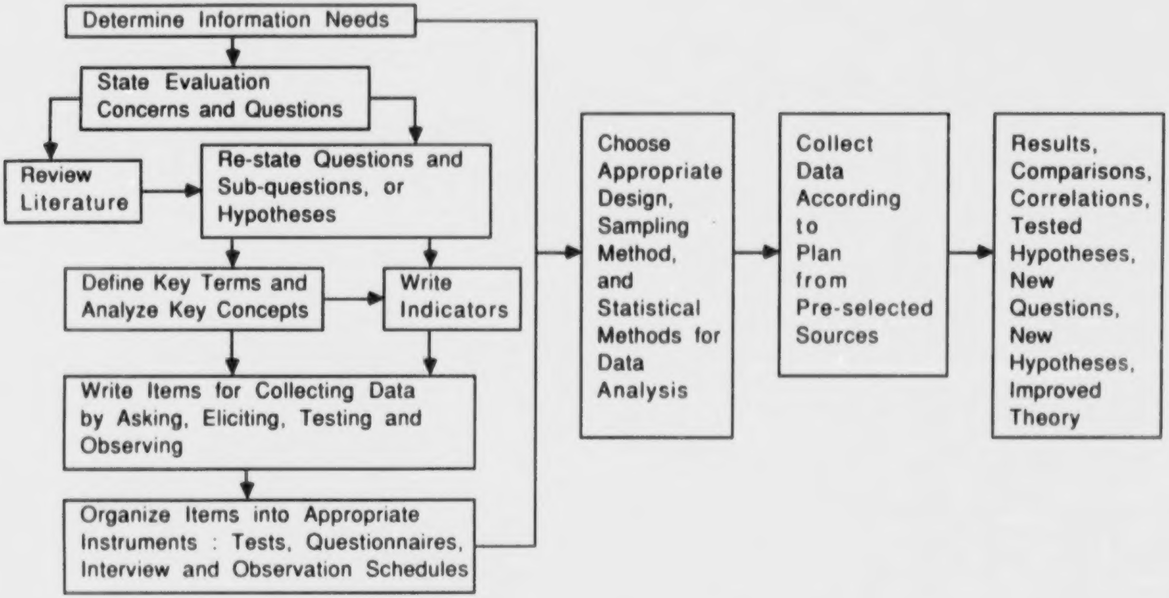

Determine Information Needs
State Evaluation Concerns and Questions
Review Literature
Re-state Questions and Sub-questions, or Hypotheses
Define Key Terms and Analyze Key Concepts
Write Indicators
Write Items for Collecting Data by Asking, Eliciting, Testing and Observing
Organize Items into Appropriate Instruments : Tests, Questionnaires, Interview and Observation Schedules
Choose Appropriate Design, Sampling Method, and Statistical Methods for Data Analysis
Collect Data According to Plan from Pre-selected Sources
Results, Comparisons, Correlations, Tested Hypotheses, New Questions, New Hypotheses, Improved Theory

Figure 8: The Process of Rationalistic Evaluation (RE)

SECTION A: Tools and Instruments

Whatever the nature of the information-gathering approach for evaluative purposes, some types of data gathering will be involved. There will have to be some seeing, observing, questioning, interviewing, eliciting, and testing. There are, of course, essential differences in how information gathering is done in NE as contrasted with RE. The instruments used in NE are unstructured or very loosely structured. There are many open-ended questions. The data are qualitative. In RE, structured instruments are considered a merit. The data are nominal or ordinal. Even when "qualitative data" are collected, they are "quantified" to be processed as numerical or nominal data. It can be easily surmised that an MIS, since it typically stores numerical data, on the surface is seen to have greater affinity with RE than with NE.

Under Chapter 7 above, Tools and Techniques of Implementing an MIS, we dealt with the topics of making and administering tests of achievement. Later, in Chapter 11, we dealt with unstructured interviews and observations as special instruments of evaluation in the naturalistic mode. Somewhat arbitrarily, we had left the discussion of structured questionnaires, structured interviews and structured observations for this part of the monograph. We now return to these structured instruments.

Structured questionnaires

As the name suggests, structured questionnaires are "structured" in regard to the questions to be asked; the exact words to be used in presenting those questions; the sequence in which those questions will be asked; and the format in which answers should be elicited and recorded. (It should be noted that questionnaires are not simply a set of "questions" with a question mark at the end. Questionnaires can include scales, multiple choice items and other devices for eliciting and recording responses.)

Structured questionnaires are often distributed by mail. In special cases these may be distributed by hand, and in fact may be filled by a field investigator. This, for instance, will be the case when such a questionnaire is used with a selected group of illiterate adults. In such an instance *a structured questionnaire becomes a structured interview*.

Questionnaires should be short and well designed. Since they will be filled independently by the respondent, they should include instructions which should be clear and easy to understand. A short introduction should provide the purpose of the questionnaire and explain how the data provided by the respondent will help the respondent and the community in general. Anonymity of the respondent should be assured and ensured.

In dealing with tests in Part III, Chapter 7 (Section C), we have suggested that tests are tests of knowledge. Questionnaires and interviews are also in a sense tests of knowledge. The difference is that these are tests of the "particular knowledge" that an individual may have and may be willing to contribute. It is not the general knowledge of the subject matter but the private knowledge of a person -- information personally available, his or her perceptions, and attitudes and opinions of various kinds. (As we have indicated elsewhere, some achievement test items may sometimes be hidden in a questionnaire.)

Local adaptations of available questionnaires

Evaluators will typically have to design their own questionnaires to suit the special social and program contexts of their evaluation studies. However, questionnaires on similar subjects developed by other evaluators elsewhere may sometimes be adapted for use. Many useful items could be borrowed from other questionnaires with very little rewriting.

Writing good questionnaires

Good questionnaires are made with clear objectives in view. They ask what the evaluator needs to know, avoiding unnecessary questions. But the important questions are not forgotten. Standard demographic information such as sex, age, occupation, income, etc., is always asked so that it is possible to interpret the overall responses received.

Item writing for questionnaires offers an additional set of problems since (1) they may ask for private knowledge that the respondents may be unwilling to part with; and (2) they may seek to elicit opinions and attitudes that the respondents may not be prepared to share honestly. Attitudes in regard to family planning, inter-marriage between people from different tribes, and taboo foods may not be honestly expressed. The respondent may supply "socially acceptable" responses. They may tell the literacy evaluators what they assume to be the proper attitude to have rather than what the respondents actually believe in regard to a particular aspect of their social or cultural world.

To solve some of these problems, writers of questionnaires may make the intent of an item less direct and may ask the same question in different ways within the one questionnaire.

Once again, pre-testing of questionnaires is important before administering them on a large scale as part of an evaluation study. Such pre-testing will bring out many problems in the questionnaire.

The list on this and the next page shows some examples of errors actually made by beginning evaluators while writing items for questionnaires. Many such problems may be caught in the process of pre-testing of the questionnaire. With practice, item writing for questionnaires will surely improve.

ITEMS	COMMENTS
A district officer is asked: After information has been communicated to the chiefs/assistant chiefs in your area, how is this acted upon?	Can the district officer really tell? Wouldn't it be better to get this information from the chiefs themselves? Aren't we asking the wrong respondents?
A community level nutrition worker is asked: What do you engage in during your home visits?	Isn't this too general a question?
A subject is asked: Do you attribute your friend's failure to laziness?	What is laziness? Do we all mean the same thing by word laziness?
A subject is asked: Do you think you were in good health during the period of the training course?	Do the subject and the evaluator understand the same thing by good health? What if the student has not been too well, but never too sick to miss classes for long? Shouldn't we ask the question in terms of days missed because of sickness?

ITEMS	COMMENTS
An extension worker under training is asked: Was your visit to the farmer useful?	Useful for whom? In what way? On the basis of what kind of evidence, using what criteria?
The headmaster of the school is asked to judge the student-teacher's commitment to work in terms of: --unsatisfactory --below average --average --above average --outstanding	How do we ensure that the evaluator and the headmaster mean the same thing by work commitment ? Do we define commitment in terms of punctuality, or carrying an overload of work, or offering tutorials to weak students? How will the headmaster come to acquire the knowledge on which these judgements will be based?
A local extension worker is asked by the evaluator: Are locally made audio-visual materials better than those	Does "elsewhere" mean in another locality? National headquarters? A commercial producer? Does "better" mean produced elsewhere? better in production values or in terms of instructional relevance?
A cooperative assistant at the community level is asked: How many of your earlier students still practice reading skills?	Wouldn't most of them say "Many"? Isn't it a loaded question?

Interviews

Interviews are used by evaluators both for rationalistic inquiry and naturalistic inquiry. In the context of the rationalistic paradigm, interviews are structured or semi-structured. By semi-structured interviews we mean basically structured interviews, with some probing questions allowed to seek further explanations.

As we have indicated before, structured questionnaires when administered in person become structured interviews. The structured interview, therefore, has the same problems and concerns of design, item writing and display of data as does the structured questionnaire. But since interviews are conducted in face-to-face situations, they pose some additional problems and challenges. The interviewee must be motivated to give the interview and to invest the time required for completing the interview. The interviewer should be able to establish trust and rapport without influencing the responses of the interviewee. In rural settings of developing countries, it may not be possible always to take the interviewee (especially the female interviewee) aside for a long private conversation. On the other hand, the interviewer should ensure that an individual interview with a young mother does not become a family interview.

Sometimes family interviews may just be the thing we want. But then we should plan and work for a family interview. The point is that an individual interview should not be confused with a family or group interview.

It is also possible to use more than one interviewer in conducting an interview. A chief in a rural community may be interviewed about his work by a full panel of interviewers.

Scales included in the questionnaires or independent scales for recording attitudes and opinions

As we have indicated above, questionnaires include not only questions. A variety of items may appear in questionnaires, including scales. In its simplest form a scale may look like that on the next page.

YES UNDECIDED NO

or AGREE DON'T AGREE DISAGREE

A scale could be made more sensitive by simply adding further ordinates, as in the following:

STRONGLY AGREE AGREE DON'T AGREE DISAGREE STRONGLY DISAGREE

The intermediate ordinates of scales do not always have to be named and may in fact be left without labels. Note the following scale that uses seven ordinates without labels, with two bipolar opposite ends:

Creative ___ ___ ___ ___ ___ ___ ___ Uncreative

These scales can be converted into multi-dimensional scales by using many bipolar dimensions such as:

creative-uncreative
hard-easy
flexible-inflexible
exciting- dull
strong-weak
scientific-artistic
objective-subjective
organized-disorganized
relevant-irrelevant
practical-impractical
active-passive
demanding-undemanding
involving-alienating
modifiable-unmodifiable
and motivating-alienating

These scales can be analyzed together for a firmer view of the attitudinal or value structure of an individual. Sometimes such scales may be given numerical values as in the following:

Organized ___ ___ ___ ___ ___ ___ ___ Disorganized

+3 +2 +1 0 -1 -2 -3

This permits quantification of data collected through scales using qualitative labels.

When applied to a group, these scales can be used to describe the structures of groups by working out the percentages of the responses. For example:

Creative 3.7 22.2 26.6 23.8 18.3 5.8 Uncreative

% % % % % %

Field observations

Field observation, again, is a data collection strategy that can be used within both the rationalistic and the naturalistic paradigms. Field observation within the rationalistic paradigm may be based on random sampling and may be highly structured. Within the naturalistic tradition, field observation will be unstructured and purposive. We may make participant observation or nonparticipant observation.

Evaluators want to make field observations to get a direct sense of the reality without an intermediary having to see and interpret it for us. Observation is not, however, a matter simply of opening our eyes and ears to people in real-life situations. We have to train our eyes and ears and must learn to record our observations. Diaries, check-lists, maps and diagrams, schedules, sociometric scales, rating scales, and cameras can all be used to record observations.

Observation schedules are by no means easy to write, and a whole range of errors can creep into them. Examine the following examples:

ITEMS	COMMENTS
Is the student-teacher audible enough to pupils sitting at the back of the class?	Can this be observed? Or do we have to ask the back-benchers about it? Or should the evaluator walk to the back of the room and listen?
Does the student-teacher speak with confidence?	What should we look for when observing a display of confidence?
What economic status do the loanees have?	Can one "observe" economic status as such?
How did the loanees use the funds they obtained from the cooperative society: married second wives, paid children's fees, engaged in heavy drinking, or bought new clothes?	How can we observe this history of behavior in a visit or during a short period of observation? Such information will have to be collected through alternative means.
An observation schedule seeks to observe: -- attitudes of people before the public meeting starts; and -- attitudes of the people during and after the public meeting.	Is it possible to observe these? Do attitudes change in the course of a public meeting? Do attitudes show on people's faces?
Does the cooperative society keep the books required under the law?	Okay, but isn't this a matter of an audit rather than observation?

Records and documents

Records and documents are important sources of data for the evaluator. The analysis of records and documents may be quantitative (suited to the rationalistic paradigm) or qualitative (suited to the naturalistic paradigm).

The ethics of buying data

The question has often been raised: Should an evaluator pay his or her respondents for participation in an evaluation study? There is no simple "Yes" or "No" answer. Knowledge production is a social function; and in the case of an evaluation study, the social use of evaluative information can often be quite clear both for evaluators and for respondents. If the evaluator is working in behalf of the government or a non-profit making voluntary agency, it is public interest which is being served by the evaluation. The respondents, as good concerned citizens, should freely participate in the evaluation study.

If, however, a subject is put in a position of having to choose between working on a construction site for the day or participating in your evaluation study, you should then pay to compensate for the wages lost by the respondent.

Things to do or think about

1. Develop a detailed list of factual statements, principles, skills, and attitudes that you want your trainees to have learned by the end of your training course.

2. Have you been interviewed recently by someone as part of an evaluation or a survey of some kind? What do you remember that was good about the interview? What did you find irritating or unacceptable? Was the interviewer able to win your trust?

3. Write an observation schedule on "Working habits in the office". Try it on a colleague. Ask your colleague to then try it on you.

SECTION B: Data Collection

Many of the problems of implementing evaluation studies have been referred to directly or indirectly in other parts of this monograph. A systematic and self-contained discussion of the practical problems of conducting evaluations may, however, be more helpful and is included below.

Circumstances are sometimes stronger than men and women are. An evaluator cannot control wind and weather, nor drought and famine. One can only cope with such circumstances and do the best possible. But many other possible sets of circumstances can be anticipated, and one should be ready for them.

A new set of collegial relationships

Evaluation is unusual business. Even when it is an evaluation of your own work by yourself, you disturb the existing relationships with your colleagues. It is important that you keep your feeling of self-importance in check and inform all concerned about what you are doing and why. Personal fears must be assuaged and professional jealousies must be relieved.

Evaluation will always make unusual demands on those who work with you in the office and in the field. The evaluator has to transform all his officers, colleagues and assistants into professional collaborators. The evaluator has to receive the blessings of those above; establish fair exchanges with those at the same level; and receive help from those below, not by ordering around but by sharing excitement as well as credit for the work done. Due acknowledgment must be made, both verbally and in writing, to those who provided advice or assistance.

Training of field investigators

In most cases, you as an evaluator will not be able to collect all data single-handedly. You will need the assistance of colleagues and other field workers. It is important that those who have been mobilized as field investigators are provided with appropriate training and orientation. The evaluator may not always want to inform the field investigators about the evaluation hypotheses or questions, in order to keep out the personal biases of the field investigators. But the field investigators must be fully trained in the requirements of administering the evaluation instruments. (I learned a simple fact the hard way: Do not fill your questionnaires or interview schedules in ink. It can wash off in the rain. Use lead pencils of ballpoints that can survive contact with water.) Such orientation and training may have to be fairly extensive if in-depth interviewing is involved.

It is important that the evaluator is able to stay in constant touch with the field investigators to answer their questions and solve unanticipated problems.

Piggybacking on existing institutional resources

It is important that literacy evaluators learn to piggyback on existing institutional resources. This is especially important in the case of transportation facilities. Travel to the field should be made to fit the travel plans of various officers from the parent department as well as other sister development departments.

Dealing with the respondents

The evaluator cannot anticipate famines and funerals, but must be aware of the seasons for migration of potential respondents, their daily patterns of work, and their festivals and holidays.

The investigator must be able to stay in the area long enough to wear off the novelty effect of his or her being there; to establish a rapport with the people; and to administer the questionnaires or to conduct the interviews. The evaluator may have to use a third person to accompany him or her to conduct interviews with young mothers who may feel embarrassed being all alone with the investigator. In such cases, the third person will have to be chosen with care and the rules of conduct during the interviewing or questioning will have to be properly understood by everyone involved.

There will be situations when respondents will expect to be paid for being subjects of an evaluation study. As we have indicated elsewhere, evaluators (and researchers) should not pay for data unless a respondent will be losing wages in cash by participating in the evaluation study.

Changes in samples and instruments

In naturalistic evaluations, sampling is purposive, and samples are developed and redefined to suit the circumstances. In so-called rationalistic evaluation, samples are pre-determined and pre-selected. It will often happen that the evaluator is not able to collect data from the pre-selected sample and is obliged to make substitutions for the respondents lost or is forced to make do with smaller samples. It is not possible within the scope of this chapter to deal with the complex issues of sample attrition and sample substitution. A general piece of advice can be offered, however. This is that evaluators must keep a precise and honest record of the changes made in the samples so that appropriate judgements can be made at the stage of interpreting data and results.

There will also be instances when changes in the evaluation instruments will be necessary. Some questions may not be understood by the respondents in an evaluation study. Some questions may be unanswerable, and some others the respondents may refuse to answer. The evaluator should be in touch with the field investigators (where field investigators are involved) to discuss problems and make the necessary changes. Changes made in the instruments should be followed uniformly by all field investigators. Clearly, some of these situations can be avoided by proper pre-testing of evaluation instruments.

Handling of completed instruments

Problems can arise from careless handling of completed instruments. Questionnaires and interview schedules can get lost or damaged in the rain. Data are precious and should be treated as such. Field investigators should be instructed clearly in regard to mailing and despatch of data. Should they be sent by hand with officials travelling from the field to the city office? Should they always be mailed? How should they be packed? Should they be sent by registered mail?

Completed questionnaires and instruments can be mixed up in the evaluator's office. These should be properly marked and coded as soon as received.

Things to do or think about

1. What are some of the problems that you anticipate in the course of data collection in your setting?

2. What are your suggestions for evaluators in regard to establishing fruitful collaborative relationships with their colleagues and subordinates?

3. Can you think of cases where problems in data collection in the field killed an evaluation study?

SECTION C: Processing and Display of Data

Data are just that -- *data*; data are not *information*. After tests and questionnaires have been administered, interviews have been conducted, field observations have been made, and records and documents have been examined, what we then have are *raw* data. Raw data, in themselves, are not information. Raw data must be coded, weighted, collated, processed, analyzed and synthesized to produce information that can be used to make program decisions.

Data processing and data analysis: Meaning and process

Elsewhere in the handbook, we have pointed out that data processing and data analysis are overlapping tasks. Data processing involves all that is involved in counting, collating, consolidating, standardizing and presenting data in particular formats to enable analysis -- analysis being a combination of the logical and analogical, the intuitive and the analytical.

Again, it should be recollected that in RE, the objectives typically are to make normative statements, to make comparisons or to test correlations. In this case then, the task of data processing is to do whatever is necessary to put data in tables and grids so that the standard statistical procedures can be applied on them as part of data analysis. Thus, the skills and techniques of data processing involve coding, scoring, weighting, standardizing, ranking, etc. Data analysis involves making normative probability statements, and running statistical tests for correlations and differences.

The essentials of the process of data processing/data analysis

Let us take a total view of the process of data processing/data analysis and try to understand the essentials of the process. This is shown graphically in Figure 9.

The essence of the process shown in the figure is to look for relationships and patterns in the data that will enable the evaluator:

1. To make a set of normative, probabilistic assertions about what is. (What is the relative importance of reasons given by adults for dropping out of the literacy program? What is the general structure of achievement in reading of adults who started in a literacy class one full year ago?)
2. To compare Knowledge-Attitudes-Performance (KAP) of groups and communities at one particular time; over a period of time; along a social hierarchy; and within differing contexts. (Which group has greater functional knowledge, Group I or Group II? Is the rate of adoption of innovations in community X better today than it was two years ago? Do administrators at an upper level of the program hierarchy have a different view of a particular phenomenon than do field workers? Does a particular method of teaching work better in the urban context as compared with the rural context?)
3. To correlate performance along one aspect with performance along other aspects. (Do adults perform better on arithmetic than on reading? Do those who have good reading scores also have equally good writing scores? Are literate adults better adopters of innovations?)

Please note that these *typical* questions will have been anticipated in the *design* of our studies and will have influenced our *choice of respondents* and sources of data and our *selection of samples*. It should be remembered also that some of the data will be converted into information while some will be used to describe the context for interpreting the information generated.

Some mechanical tools and routines of data processing/data analysis

We have all heard stories of how some beginning evaluators are overwhelmed by the data they have collected. They do not know

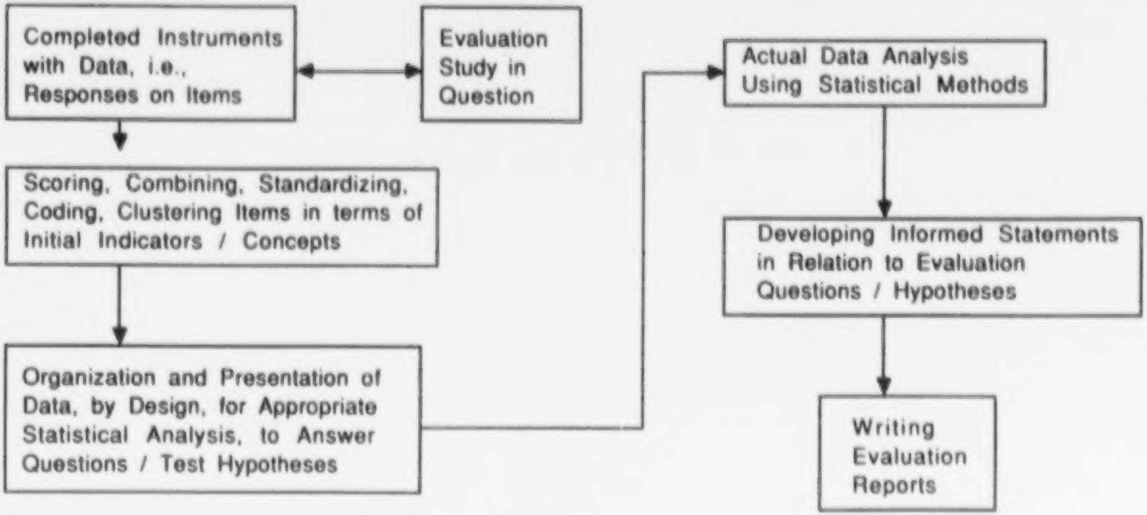

Completed Instruments with Data, i.e., Responses on Items
Evaluation Study in Question
Scoring, Combining, Standardizing, Coding, Clustering Items in terms of Initial Indicators / Concepts
Organization and Presentation of Data, by Design, for Appropriate Statistical Analysis, to Answer Questions / Test Hypotheses
Actual Data Analysis Using Statistical Methods
Developing Informed Statements in Relation to Evaluation Questions / Hypotheses
Writing Evaluation Reports

Figure 9: RE - Focus on Data Processing and Analysis

what to do with the bundles and bundles of questionnaires and interview schedules they have got filled. Some end up reading through some or all of their data, taking notes, and writing personal and impressionistic essays on their experience of doing field work, and stating what they have learned in general. In the following are some of the mechanical tools and routines of data processing/data analysis that you should find helpful in coping with the data you have yourself collected:

1. A good supply of ruled and plain paper
2. A supply of lead pencils and a pencil sharpener
3. Erasers
4. If possible, a bottle of correction fluid
5. A pair of scissors
6. Scotch tape with dispenser
7. Paper clips and pins, and
8. A set of colored pencils

Another basic suggestion

In the process of data processing and analysis, write only on one side of the paper. Use a separate sheet of paper for each single idea or table that you develop. This will help you later in trying different organizations of the material. You do not have to use nice fresh paper for this stage of data processing. You should use discards from cyclostyled materials and any other scrap paper you can get hold of. For making tables by hand, use ruled paper so that rows of data can be read without confusion. Be careful about the spacing of numbers in columns:

125	125
5	5
11	11
is *correct*	is *not* *correct*

Do not write over your own writing. Use an eraser; or strike out and write afresh. Otherwise you will yourself wonder later whether you had changed a 3 into an 8 or an 8 into a 3.

Clustering and identification of data pieces

For the sake of convenience, let us give the name *data pieces* to all the individual tests, interview schedules, observation schedules and questionnaires filled and returned by investigators and respondents. The very first thing to do when all the data pieces are in, will be to arrange and identify the various pieces by *assigning them numbers*. Different clustering arrangements will be appropriate in different cases. Where respondents are not anonymous, data pieces may be arranged alphabetically. Other arrangements may be used to reflect clusters of data pieces by sex, age, religion or ethnicity; by training course, batch or year; by region, province or district; by literacy teacher in charge; in terms of trained versus untrained groups; and by the training methodology used.

Examine the "Super Table" included later in this section. The clustering used in the Super Table should be anticipated in assigning numbers to the various data pieces.

Such clustered organization and identification of data pieces helps at the later stages of data analysis. Once organized according to need, all data pieces should be given permanent numbers in the upper right-hand corner on the face of each piece. Color coding may also be used to help quick recognition of various data pieces.

If a whole set of instruments -- an achievement test, an interview schedule, an observation schedule and a questionnaire -- have all been used with the same one group of respondents, then a matching numbering system should be used. For example:

Name	Interview Score (I)	Observation Score(O)	Questionnaire Score (Q)	Test Score (T)
Abram	I-1	O-1	Q-1	T-1
Binti	I-2	O-2	Q-2	T-2
Camaro	I-3	O-3	Q-3	T-3
Daudi	I-4	O-4	---	T-4
Elice	---	O-5	Q-5	T-5
Fakouri	I-6	O-6	Q-6	T-6

Make sure that you write I and T and O and Q clearly enough so that I is not confused with T, and Q is not confused with O.

Note that in the above display, Daudi's questionnaire is missing, as is Elice's interview schedule. However, Fakouri still gets numbers I-6, O-6, Q-6 and T-6 for his data pieces. In other words, all data pieces for the same one person are given matching numbers.

The need for immersion in the data

After the data pieces have been arranged and numbered, it will be time to do two further things: to recollect the evaluation questions that needed to be answered by the evaluation study; and to become "immersed" in the data already collected.

Write out the list of questions you want the data to answer. If there are sub-questions to the questions, write them out also. For an example, read the following set of questions:

1. How are *trained* assistant adult education officers different in regard to their overall performance from *untrained* assistant adult education officers?
    - 1.1 How do they differ in regard to their technical knowledge about development and adult education?
    - 1.2 How do they differ in regard to their knowledge of the literacy methodology being used in the program?
    - 1.3 How do they differ in regard to their supervision styles, and diagnostic and problem-solving skills?
    - 1.4 How do they differ in terms of their attitudinal orientation to adult learners, rural communities and their own work?

Remember that these would have been your guiding questions when you began the evaluation study. But changes are often necessary as questions are reformulated at the stage of data processing/data analysis.

Armed with such a list of questions, it is time to begin the immersion in the data. By immersion in the data, we mean going carefully through all the data pieces, piece by piece, page by page, item by item; studying all the responses; and making careful, written notes. You should take note of the expected, of the *un*expected and of the curious; of the emergent pattern and of the seeming relationship, as you go through the data. This immersion may require more than one dip; that is, you may have to do more than one reading of the data pieces. The time used in going through this process is always well spent. Therefore, be patient.

Possibilities and limitations of the data collected

This will also be the time to discover the unanticipated possibilities of the data. For example, a questionnaire used with students of agricultural extension to evaluate their attachment experience, may be full of information about prevalent practices on butchering meat animals; or on the popularity of poultry farming in a particular region. On the other hand, serious problems may be discovered with the data during the immersion process. Some questions in the test may have been consistently misunderstood. Other questions may have received "socially acceptable" responses, not the real answers.

Some pieces may have to be discarded altogether for being incomplete or dishonest. It may become clear to the evaluator that available data will not make an overwhelming case for or against a particular position or approach; and the evaluator may have to warn readers against drawing unwarranted conclusions. All this should be taken note of, in writing, during the process of immersion in the data.

The Super Table

Data processing by computer is a different question altogether. But if data processing/data analysis has to be done manually, with paper and pencil, as most of you will be doing, then the best thing is to prepare a "Super Table" on which ideally *all* the data relevant to one major evaluation question (and sometimes a whole evaluation study) could be accommodated, in rows and columns, and appropriate clusters for *one total look*.

It is amazing how much can be put into the same Super Table (affectionately called "the Blanket", by the participants of the evaluation workshops in Kenya). An example is given on pages 240-241.

The various columns of the Super Table can be used to include scores on a variety of aspects of KAP. Time can also be reflected in the columns. For example, [a] could be scores before teaching and [b] after teaching. Scores in column [c] could be innovation adoption before the program began and under column [d] after the programs had been in effect.

SUPER TABLE (THE BLANKET) ON A FUNCTIONAL LITERACY PROGRAM

	COLUMNS							
	[a]	[b]	[c]	[d]	[e]	[f]	[g]	....[n]
Region X								
Method 1								
Teachers (Trained/Male)								
Learners:								
Males 0-15 Years								
M1								
M2								
M3								
M4								
Males 16-45 Years								
M5								
M6								
Males 46-65 Years								
M7								
M8								
M9								
M10								
Females 0-15 Years								
F1								
F2								
F3								
F4								
Females 16-45 Years								
F5								
F6								

	[a]	[b]	[c]	[d]	[e]	[f]	[g]	....[n]
Females 46-65 Years								
F7								
F8								
F9								
F10								
Teachers (Untrained/Male): List Male and Female Learners separately in appropriate age sets.								
Teachers (Trained/Female): List Male and Female Learners separately in appropriate age sets.								
Teachers (Untrained/Female): List Male and Female Learners separately in appropriate age sets.								
Method 2								
Repeat for different categories of teacher (Male and Female, and Trained and Untrained), separating learners by sex and age sets.								
Region Y								
Repeat for different methods (Method 1, and Method 2), teacher categories, separating learners by sex and age.								

Fitting data in the Super Table
The type of Super Table we are proposing is not good for words and phrases. It is best for numbers (5, 7, 11, 21, 51, 101); for letters (A, B, C or D); or for marks (or X). In other words, before we can prepare Super Tables, we must learn to score, codify, weight, standardize, and rank order data.

Coding. Coding means to assign a particular code to a particular category of response. The following are examples of coding frames:

Code 1	Prefers condoms as family planning aids	A
	Prefers an IUD for his wife	B
	Prefers to do family planning by abstinence	C
Code 2	Has insufficient (low) nutrition information	L
	Has average (medium) nutrition information	M
	Has high degree of nutrition information	H

Scoring. Scoring is assigning numerical values to particular responses or to particular levels of performance. Attitudinal responses will often be qualitative and will need to be scored. The same is true of performance scores which may involve observation of performance, judgements on what is observed, and the change of judgement into some sort of quantitative score.

Weighting and combining scores. As teachers we know that in writing achievement tests, we can assign different marks to questions on the question paper, depending upon the difficulty or the importance of particular questions. This differential allocation of marks to different questions (and answers) is called weighting. Weighting is also involved in the analysis of opinion and attitude questionnaires and observation schedules. Needless to say, allocation of weights to responses on an attitudinal scale should be undertaken with care, especially in regard to the values of neutral, positive and negative responses.

Standardizing. To standardize scores is to so treat them that they can be compared using the same yardstick. Getting 13 marks out of 20, is better than getting 14 marks out of 25. A profit of 75 shillings on a 400 shilling investment is not easily comparable with a profit of 15 shillings on a 50 shilling investment. When both profits are standardized as percentages (18.75% versus 30%) they are easily comparable. Working out percentages is an important way of standardizing scores.

Evaluators of literacy and development programs will often have to compare scores made by individual trainees on a variety of achievement and performance tests. Each time scores are to be compared, the evaluator should check if prior standardization of the scores will be necessary.

Ranking. Ranking has common-sense meanings. It means simply to put the scores of achievement or performance in a sequence so that the highest score comes first and the lowest score comes last. (The arrangement could be the exact reverse, giving the lowest score the first position and the highest score the last position). Where more than one respondent has the same score, the tie is broken as follows:

SET A		SET B	
Scores	Ranks	Scores	Ranks
69	1	69	1
65	2.5	65	3
65	2.5	65	3
61	4	65	3
		61	5

In other words, the tied scores are each ranked to be in the middle of the untied rank positions: in the first example 2.5 is in the middle of 2 and 3; and in the second example 3 is in the middle of ranks 2, 3 and 4.

In the following examples the techniques of coding, scoring, weighting, standardizing and ranking have been demonstrated.

EXAMPLE 1

In evaluating the effectiveness of a training program for teachers of agriculture, a classroom observation schedule used the following items:

Teaching Skills

(i) Provides introduction to the lesson	Right	Wrong	Confusing
(ii) Changes method according to need	Yes	No	Reluctantly
(iii) Helps students recapitulate the lesson	Periodically	Not at all	Only at the end
(iv) Accepts and answers questions	Always	Not at all	Sometimes
(v) Gives individual attention	To all	To none	To some poor students
(vi) Helps the students write notes	Always	Not at all	Sometimes

---

We do not wish to make any comments here on the merits or demerits of the items as written. The point we want to make here is simply that some numerical values must be assigned to the judgements made during the observation; and that those values must be aggregated for use in data analysis. For example, approved behavior may be assigned a score of +1, an indifferent behavior may be assigned the value of 0, while an unacceptable behavior (which will hinder learning) may be assigned a value of -1. This will enable the evaluator to come up with an aggregated score for the teaching skills evaluated in Example 1, as suggested in the following illustration:

(i)	+1
(ii)	-1
(iii)	+1
(iv)	0
(v)	0
(vi)	+1
Total score:	2

It is important to note that different types of question can be asked from the same data. For example, consider the question: Do student-teachers, typically, help children to recapitulate ideas given in a lesson? Looking at answers on item (iii) above, for all the student-teachers tested, an answer to this question can be found.

EXAMPLE 2

A questionnaire (filled by each student individually, but sitting as a group in a large hall) sought to evaluate the effectiveness of field attachment of agricultural students. The part of the questionnaire dealing with disease control had been broken down into the following items:

1. What notifiable disease(s) did you come across?
2. What methods of control and prevention were used?
3. Mention the vaccination campaigns you saw.
4. Enumerate the diseases against which vaccination was done.
5. What were the reasons for vaccination?
6. How was the vaccination organized and carried out?
7. How was the vaccine administered? Indicate any special precautions taken.
8. How many animals were vaccinated?
9. What was the dosage of the vaccine?
10. What was the cost of the vaccine per dose?
11. How did the farmer pay for it?
12. What is the duration of immunity for the vaccines used?
13. What was the type of vaccine used -- live, attenuated or dead?
14. What were the problems encountered in the vaccination campaign?
15. How were the vaccines used handled?

As we can see, these questions are a combination of (i) knowledge by the student of technical information; (ii) recall of "what" was done and "why" in some problematic situation in the field; (iii) information about some local happenings during the period of the student's attachment; and (iv) descriptions of professional actions and technical practices seen by the student during the field attachment but over which the student might have had no control.

In this case, the evaluator will first have to separate items of student's responsibility from those items which were part of the context; and then will have to make judgements about the quality of student performance *in the given circumstances*. The evaluator may assign A, B, C, D and E grades (or some number grades) to the performance of each student.

Once again, we should note that many different uses can be made of this data, in addition to evaluating student performance. Using the same data, one could develop evaluations of dip management or clinical centers in the country; learn about the diffusion of new skills within rural communities; or learn about farm management practices, in general.

The problems of scoring interview and observation data, to change qualitative into some kind of quantitative data, cannot be completely eliminated in this value-laden world of ours. However, some serious problems can be mitigated at the stage of instrument design and item construction. Tools, and items included in those tools, can be so designed as to elicit answers that are more easily amenable to quantification.

From the super table to summary tables

After the coding, scoring, standardizing and weighting have been completed, it is time for the evaluator to have a full and complete, overall look at the data. This, as we have suggested above, can be done by developing super tables that show at one glance the responses made by all the different subjects on a total test, a whole questionnaire or some other instrument in the context of an evaluation question (or study). These super tables may be as large as the size of your working table, covering a large part of your office.

A careful look at a super table or blanket would suggest many different leads to the evaluator in regard to response patterns, and differences and correlations between items. By focusing on the various rows and columns of the larger blanket, one can develop many useful crossbreaks and summary tables that answer particular evaluation questions.

Statements on what is happening

One simple summary table that could come out of the Super Table would be about reasons for dropout, with frequencies.

Suppose that under one of the columns (say, column [f]), we reported the status of participation (A = Active; D = Dropout) as well as the reasons for dropping out (D-1; D-2; D-3; D-4, etc.) The reason code could have been: 1 = Sickness of self; 2 = Sickness in the family; 3 = Moving away for economic reasons; 4 = Lack of interest in program objectives and content, etc., etc. These codes could now be converted back into their qualitative descriptions and could be shown as percentages as follows:

SUMMARY TABLE 1

Reasons, with Frequencies, for Learner Dropouts

Reason for Dropout	Percentage of time mentioned
1. Sickness of self	16
2. Sickness in the family	17
3. Moving away for economic reasons	21
4. Lack of interest in program objectives and content	7
5. Interpersonal problems with teachers or with other learners	19
6. Drunkenness	13
7. Feels learning objectives have been fulfilled	7

Please note that the data in the above summary table have been made up by way of demonstration.

Comparisons and correlations between categories of score

Summary tables involving comparisons can be of many kinds: between male and female groups; across age distributions; between rural and urban regions; between different instructional methods; between trained or untrained teachers; in before and after format; and across time periods. Lessons in a primer could be analyzed or the items in a test could be tested for reliability and validity. Correlations could be established between performance scores. Some illustrations are given in the following Summary Tables 2-10.

SUMMARY TABLE 2

Averages of KAP Scores by Sex

	Literacy Skills			Attitudes	Performance
	R	W	A		
Males					
Females					

R = Reading Score; W = Writing Score; and A = Arithmetic Score

SUMMARY TABLE 3

Averages of KAP Scores by Sex and Age

	Literacy Skills			Attitudes	Performance
	R	W	A		
Males					
0-15 Years					
16-45 Years					
46-65 Years					
Females					
0-15 Years					
16-45 Years					
46-65 Years					

SUMMARY TABLE 4

Learner Achievement by Categories of Teacher

	Literacy Skills			Attitudes	Performance
	R	W	A		
Male Teachers					
Trained					
Untrained					
Female Teachers					
Trained					
Untrained					

SUMMARY TABLE 5

Achievement by Regions and Methods Used

	Literacy Skills			Attitudes	Performance
	R	W	A		
Region X					
Method 1					
Method 2					
Region Y					
Method 1					
Method 2					

SUMMARY TABLE 6

A "before" and "after" evaluation design may now appear as a model of data analysis as follows:

Improvement in Nutrition Information and Behavior after an Instructional Intervention

	Knowledge of nutrition	Relevant nutrition-related behavior
Before the introduction of the course		
After the introduction of the course		

SUMMARY TABLE 7

Improvement in KAP Scores of a Group of Learners Over Time

	Before Intervention	During Intervention	After Intervention	Later as Retention
Knowledge				
Male				
Female				
Attitudes				
Male				
Female				
Performance				
Male				
Female				

In the crossbreak shown in Summary Table 8, all learners have done almost equally well in Lesson I on item 1, and equally poorly on item 2. Under items 3 and 4 no patterns seem to emerge. Maybe items 1 and 2 are not good items since they do not help us separate good students from bad ones. Or, maybe items 3 and 4 are poorly written and need to be reworked. Why is it that all learners do poorly on items on lesson II, but rally in lesson III? Is it that lesson II is unduly difficult? All such questions can be raised by such data.

SUMMARY TABLE 8

Scores of learners could also be used to test items included in a test or for pre-testing a set of instructional materials.

Testing Test Items or Testing Instructional Materials

Lesson	I					II					III				
Question Number	1	2	3	4	Total	1	2	3	4	Total	1	2	3	4	Total
Total Possible Points	5	5	5	5	20	4	5	6	5	20	6	5	3	6	20
Learner A	5	3	2	1	11	3	2	2	1	8	4	4	3	5	16
Learner B	5	3	4	4	16	4	0	5	4	13	5	5	3	4	17
Learner C	4	2	5	4	15	4	4	5	5	18	6	4	3	5	18
Learner D	5	3	1	2	11	2	3	2	1	8	4	4	3	5	16

Data could also be summarized for working out correlations between different components of the program or aspects of learner performance. This is seen in Summary Table 9.

SUMMARY TABLE 9

Relationship between Attendance and Final Grade

Lesson	I			II			III		
Attendance	High	Medium	Low	High	Medium	Low	High	Medium	Low
Learner A	11				8				16
Learner B		14				13		17	
Learner C			15			18			18
Learner D	11			7					16

In the crossbreak above, visual inspection of data can give some ideas about the relationship between attendance and final grades. If no pattern seems to emerge, different additional questions can be raised.

Finally, data from the Super Table can be developed to work out rank correlations.

SUMMARY TABLE 10

Scores Arranged for Computation of Correlation

Learner	Reading Score	Writing Score	Rank in Reading	Rank in Writing
-	-	-	-	-
-	-	-	-	-
-	-	-	-	-

SECTION D: Statistical Analysis of Data

Data analysis in RE typically means statistical analysis. For example, for comparisons between independent samples (two groups from two different villages) and non-independent samples (before and after scores of the same group of learners) appropriate versions of the t-test may be used. When mean scores are not available, but proportions and frequencies are, Chi-square tests may be preferred to demonstrate differences between groups of participants and non-participants of a literacy program. Finally, for studying correlations, rank order correlations may be worked out.

The procedures of data collation and processing discussed above will enable evaluators to put their data in such forms that various statistical tests can be performed on the data. For statistical formulas, and steps in their applications, evaluators should refer to any standard textbook on statistics.

Discussion of results

In the following, some general suggestions are made about discussion of results obtained from data analysis:

1. *Relating with preconditions and entry behaviors.*
As part of the discussion of results, reexamine available data on entry behaviors and study the preconditions that prevailed when the change episode of your interest began. The phenomenon of high dropout rates from a college course, for example, may be explained

better in terms of faulty recruitment methods than by what is taught during training. The failure of a family planning program may be explained better in terms of the precondition of extremely high infant mortality rates in the region.

2. *Putting things in context.*
Analyze findings in terms of the insti-tutional and the social contexts of training programs. Do some institutional policies actually go against policies of rehabilitation of distressed families or against increasing individual savings? Does the social context promote or inhibit cooperative behavior?

3. *Relating with what is already known.*
Compare and contrast what your data tells you with what is already known. Do your findings surprise you? Are your findings reinforced by what other evaluators have found in other settings? What was expected? What is unexpected?

4. *Looking for correlations and causations.*
Data analysis will typically involve search for correlations and even for causal links. In so doing, think of the rival hypothesis -- an alternative explanation for what you see. Consider all possibilities before making broad assertions.

5. *Reexamine your assumptions.*
It is important to keep on thinking about the assumptions on the basis of which the evaluation study was designed and the evaluation questions were raised. Did those assumptions hold up? How have those assumptions changed?

6. *Relating to the limitations of data.*
Discuss results in terms of the limitations of data discovered, as the evaluation design was implemented and evaluation tools and instruments were actually used. Some limitations of data may indeed be fatal to the study and to the conclusions drawn from it. Another set of limitations may be less severe, but may introduce the need for a high degree of caution in interpreting results of an evaluation study.

7. *Setting up norms for success and failure.*
The evaluator must establish norms for success or failure of a program being evaluated. What kinds of result will provide the cause for satisfaction? What results will be interpreted as failure?

Things to do or think about

1. How are evaluation designs different from models (or plans) for data analysis?
2. List some problems that you may have come across in assigning values to responses on attitudinal scales.

CHAPTER 16

WRITING REPORTS ON RATIONALISTIC EVALUATIONS AND PROMOTING UTILIZATION OF RESULTS

The final report of an evaluation study serves many purposes. It records the history of the program as well as of the evaluation study, stating formally and clearly the findings of the study. In presenting reasons for choices of particular designs and samples, instruments actually used, and summaries of data in tables and through other display mechanisms, the final report of the study enables readers to make judgements of their own about the goodness of the evaluation study and about the reasonableness of the recommendations made. Experienced readers could draw different conclusions or additional conclusions from the data presented. A formal report, when properly distributed, expands the use made of an evaluation study.

An evaluation study, if it is to be of greatest value, must end in a written report. A written report serves at least two purposes. First, it provides an opportunity to the evaluator to organize the data collected, to systematize thinking, to draw conclusions, and to weigh and consider the implications of the study as well as its limitations. Second, the evaluation report serves as the instrument of communication between and among professional colleagues and others interested in the same or similar problems and issues.

Evaluation studies have quite often been published; and, sometimes, have brought high professional rewards to evaluators. However, publication and rewards of fame and fortune are not the right expectations to have when writing an evaluation report. These rewards may come, but one should not strain to get them every time one sits down to write an evaluation report. It is much more realistic to think in terms of making a few copies of the evaluation report to be shared, first and foremost, with colleagues in the program who should know what your evaluation study has found; who can discuss your conclusions and suggestions with you; and who, perhaps, can use the report to improve their performance within the setting of the institution to which you all belong.

We like to make a distinction here between a basic professional report and other written or oral presentations. The evaluator should prepare one basic and comprehensive report on the evaluation study. This basic report should then be used to make different written and oral presentations to different groups of people who may be interested, among them, policy-makers and planners, politicians, extension workers, and even farmers and housewives, who are often the subjects of our developmental efforts.

The essential objective of the basic report and its parts

The essential objective of an evaluation report is to make a complete record of an evaluation experience, including the background and the context of the evaluation questions; the assumptions made in posing the question(s); the evaluation design and tools used in data collection; the results obtained; conclusions drawn; and practical implications developed from the conclusions of the evaluation study. In other words, the evaluation report is a sort of mirror image of an evaluation proposal, as was discussed in Chapter 14, "Writing a Proposal for an Evaluation Study in the Rationalistic Mode".

An evaluation report, however, is more than an evaluation proposal written in the past tense. A good evaluation report includes all the information necessary for a reader to be able *to evaluate the evaluation study* itself. That is, the reader should know exactly what was done and how; using what samples and what questions; and what data was actually collected. The reader must also be told of the structure of the argument used in data collation and analysis, and what conclusions were drawn and why. In all cases, the reader should thus be able to see the strengths of the study as well as its limitations; and, where necessary, the reader should be able to do a "secondary analysis" of the data on his or her own to draw independent and even alternative conclusions. This means that actual tools and instruments, and any specimens of stimulus materials used in the study, should become part of the report as "Appendices".

This does not mean, however, that *all* raw data should become part of the report or should be put in the appendices. A report is not a device for storing and filing all the raw data that was collected for an evaluation study. Data included in the report or in the appendices should be in a collated form, already organized into tables and displays. In some cases, it may be necessary to present data in sufficiently "disaggregated" form so that it is possible for the reader to aggregate data in different ways to test assumptions and conclusions of the original evaluator; and, as we have suggested above, to draw alternative conclusions.

An evaluation report should typically (but may not always) have the parts and sections discussed below. By way of example, we have taken the case of an evaluation study in the area of development training, that is, a study to evaluate the effectiveness of a training program for development workers.

The title page

The title page of the report should show the title of the evaluation study, the name of the evaluator(s), the institutional affiliation of the evaluator(s), and the date when the report was issued.

The title given to the report should faithfully reflect the purpose and scope of the evaluation study. This same exact title should then be used throughout the study without arbitrary variations. In some cases, it may be useful to have both a long title and a short title for the same study. Once chosen, these titles should be used in other parts of the report without change. The date of issue of the report should be shown on the title page, as we have suggested earlier, but somewhere in the body of the report one should also indicate the dates and period of time during which data was actually collected. (It is possible to collect data in the first six months of 1982 and publish a report in 1984.)

The abstract

A one- to two-page abstract (that is, of about 500 words) should precede the evaluation report. This should be a complete summary and must include information about the evaluation question; samples and procedures used; and findings and their program implications. A person who does not read the full report should yet be able to get a fairly good idea of the contents of the study from reading this abstract.

General background

The first part in the main body of the evaluation report should be the general background of the study. This material will not have to be written anew, but should be adapted from the evaluation proposal written earlier. Put training for development in a larger perspective of human resource development for social change. Comment on the need for evaluation of training, in general. Be brief. No more than a page or two should be utilized.

Focus on your development sector and institution

Focus should then shift to your specific development sector such as agriculture, cooperatives, health extension, nutrition or family planning, and to your institution. Talk, for instance, about the role of your training institution and its contribution to the training of manpower needed for development. Once, again, brevity is important. One or two pages of tightly written material should be enough.

The training model in use

Present the bare bones of your model of training. Answer questions such as: What are the assumptions made about the change process in the training model in use? What are the assumptions made about the change agent's role? What are the objectives of training? What are the special training methods used? What are the K-A-P (Knowledge-Attitude-Performance) claims being made in behalf of the training program? (All these questions will not have to be answered in each and every evaluation report. Nor will these questions be answered in the order in which they have been listed here. These are the questions to "think with" as evaluators sit down to write their final reports.)

The evaluation questions asked

The evaluation questions asked in the evaluation study should be carefully listed. This list of questions must later be used in the collation and analysis of data. These questions should have two linkages. One, they should relate to the training model in use, discussed earlier. It should be clear how the training model in use generated that set of questions. Two, the questions should be linked with the subsequent organization of data in a later section and should provide the organizing principles for data analysis.

Why was this feedback necessary?

This is in fact a justification for the choice of particular evaluation questions from a whole array of possible questions generated by the training model and the institutional needs for feedback. The material from the earlier evaluation proposal on "justification" and "significance" should be used to develop this section.

Assumptions made

Assumptions made about the general change and training processes, and about the specific institutional and field settings of your evaluation study, may be stated here, as relevant. Some of these assumptions will have been stated in the earlier proposals. Some others may have been uncovered during the process of implementation.

Procedures and methods used

This section should include the general evaluation design; analyses of concepts used and special definitions assigned to terms; indicators used and the process of their development and choice; criteria to be used for evaluating success or failure of the program; samples chosen (and those originally intended); tools and instruments used (which must be placed in the appendix); field work procedures followed, including recruitment and training of investigators, and time and duration of the field work phase.

Evaluation design

Go back to the evaluation proposal and reproduce, with adaptations if any were made, the evaluation design used in the study. This model should now be congruent with the chosen model of data analysis. (See the discussion on models of data analysis in Chapter 15, Section D of this monograph.)

Conceptual analysis and definitions

You may have undertaken conceptual analyses of some concepts such as humanism and self-reliance; or may have given special definitions of your own to such words as dropout, literate, etc. These should be included in this section.

Indicators -- their development and choice

The process used in going from larger categories such as self-reliance to subcategories of larger concepts, and, finally, to the choice of indicators which can be measured, should be clarified in this section of the evaluation report.

Standards and criteria of success

It is important to indicate in the report the levels of expectation and standards set for judging success or failure of the program being evaluated. The reader should have an idea about whether to be satisfied or dissatisfied with the 30 per cent dropout rate from a literacy class or the 10 per cent rate of success in the rehabilitation of the handicapped.

Samples and units of response

Explain sampling procedures. Define the samples that were originally selected and then those that were actually used. Who were the respondents? Was it the housewife, or was it anyone else (husband, an older child), speaking in behalf of the family? Was it the chairman of the committee being interviewed, or was it anyone in the committee (or more than one person taking turns), speaking in behalf of the committee? What was intended? What actually happened?

Tools and instruments

The variety of tools and instruments used should be indicated and their choice justified. Any special procedures used in developing and pre-testing tools should be given. Changes made in tools and instruments on the basis of pretesting should be highlighted. The tools and instruments should be included in the appendices.

Field work

This section should clarify any strategy implicit in the field work -- coping with distances, or with weather conditions; piggybacking on existing systems of transportation and supervision, etc. It should, additionally, include a description of field work procedures and experience. Were investigators used? How were they trained? How were they supervised? How was communication between the evaluator and the field investigators maintained? Was there a small pilot study conducted before the final study? Did some data have to be collected twice? What was the time and duration of the field work? Was it found necessary to use a follow-up questionnaire or interview to supplement the original data?

Limitations and breakdowns

This section should look backward to field work experience, and forward to the section on data analysis and should indicate any breakdown that occurred in field work and any limitations that became apparent in data collation and analysis later.

Recording of findings

This section is the heart of any evaluation report. It has to present all relevant data in aggregated form, in effective displays of tables, charts and lists to serve as evidence for all answers given, comments offered and conclusions drawn. The list of questions drawn up earlier and the model of analysis discussed before should be used to organize the collation of data, its display, analysis and interpretation.

A separate section may deal with questions not originally asked but which the available data was able to answer.

Discussion of results

The findings must be discussed in regard to the implications for action, and guidelines for future training design. The evaluation results obtained must be discussed in terms of expectations, standards and norms. These should also be discussed in terms of the strength of data, correlations and possible causal links.

Further evaluation and feedback needs

Karl Popper has said that our knowledge and ignorance increase together! A successful evaluation study, by creating new information, might also tell us what we are ignorant of, or need to know more about. New feedback needs should be identified and suggestions in regard to further evaluation studies should be made.

Bibliography

Make a list of books, documents and government reports used in the implementation of the evaluation study and in writing the report. In the case of official documents, indicate whether they are available to the public, and if so, where they may be obtained or consulted.

Appendices

The following kinds of item should go in the appendices: copies of tools and instruments; specimens and exhibits where appropriate; collated data not used in the body of the report but of interest to readers and evaluators; lists of names of people, and institutions that cooperated with the evaluator(s) in the conduct of the study; field work schedules, maps, etc.

Example of evaluation report content

An example of a list of contents of an evaluation report is given on pages 265-268.

Reports to non-specialists: Written and oral reports

As we have suggested earlier, literacy evaluators should begin by writing one basic professional report. This basic report should then be used to write short written and oral reports for the non-specialist. These reports should be written to suit the special interests of the audience to whom the report is addressed. Oral reports should be made both informative and interesting, using appropriate audio-visual aids. These written and oral reports to special audiences should in fact become part of the process of dissemination of evaluation results.

EXAMPLE
The functioning and effects of the Kenyan literacy programme

I. Research design and implementation
  A. Background
  B. Objectives of the study
  C. Research design
  D. Research implementation
    1. Selection of locations
    2. Selection of interviewees
      a. The literacy learners
      b. The control group
    3. Preparation of research instruments
      a. The learners' questionnaire
      b. The teachers' questionnaire
      c. The literacy test
    4. Data collections
    5. Data analysis

II. The Kenyan literacy programme
  A. The national context
  B. The national literacy programme
  C. Literacy work before 1979
    1. The national literacy programme of 1979

III. The location profiles
  A. Geographical features
  B. Population
  C. Economic activities
  D. Socio-cultural characteristics
  E. Socio-economic services
  F. Self-help and local development
  G. School education
  H. The literacy program
  I. Summary

IV. The literacy centres: characteristics and functioning
  A. Buildings used for literacy teaching
    1. Original use
    2. Materials used for the construction of literacy classrooms
      a. Conditions of the teaching environments

- B. Teacher characteristics
  - 1. Categories
  - 2. Sex of teachers
  - 3. Age and marital status of teachers
  - 4. Teachers' educational qualifications
  - 5. Teachers' experience
  - 6. Previous occupation of teachers
  - 7. Second occupation
  - 8. Ties with the local community
  - 9. Teachers' attitudes and job satisfaction
- C. Teaching/learning aids
  - 1. Teaching aids
  - 2. Learning aids
  - 3. Records
- D. Learning exposure
- E. Content
  - 1. Class projects
  - 2. Guest lecturers
- F. Centre committees
- G. Average attendance
- H. Summary

V. General characteristics of the literacy learners
- A. Sex
- B. Age
- C. Marital status
- D. Number of children
- E. Languages spoken by the learners
  - 1. Mother tongue
  - 2. Ability to speak Kiswahili and English
- F. Learners' religion
- G. Learners' occupations
- H. Summary

VI. Learners' home environments
- A. Type of housing
- B. Possession of books and magazines
- C. Availability of audio-visual equipment
- D. Exposure to the mass media
- E. Summary

VII. Educational experiences of learners
- A. Exposure to primary schooling
- B. Literacy classes before obtaining proficiency certificates
- C. Reasons for joining the literacy programme and benefit seen to be derived from the programme
- D. Duration and regularity of literacy class attendance
- E. Learning after the literacy certificate
  - 1. What adults would like to learn after the literacy certificate
  - 2. Duration of literacy class attendance after the certificate
  - 3. Participation in other types of course
  - 4. Listening to the special DAI radio program
- F. Summary

VIII. Using literacy and numeracy skills
- A. Differences between locations
  - 1. Reading
  - 2. Writing
  - 3. Calculating
- B. Other types of difference
  - 1. Gender
  - 2. Speaking Kiswahili
- C. Summary

IX. Functional knowledge, attitudes and practices
- A. Measuring functional knowledge, attitudes and practices
- B. Differences between the literates and illiterates
  - 1. Knowledge
  - 2. Attitudes
  - 3. Behaviour
- C. Differences among the literates
  - 1. Location
  - 2. Other differences
    - a. Gender
    - b. Age
    - c. Year of certificate
    - d. Primary school attendance and literacy class experience
- D. Summary

X. Literacy and numeracy skills acquired
   A. Grading test results and setting performance standards
      1. Numeracy
      2. Reading
      3. Writing
   B. Results for the whole sample
      1. Numeracy
      2. Reading
      3. Writing
      4. Relations between the three types of skills
      5. Global results
   C. Differences between locations
   D. Other differences
      1. Gender
      2. Learning experience
      3. Year of literacy certificate
   E. Summary

XI. Conclusions
   A. The learners and their motivation
   B. The functioning of the literacy programme
   C. The effects of the literacy programme

Developed from Carron, G., Mwizia, K. and Tzigha, G., *The functioning and effects of the Kenyan literacy programme*. IIEP Research Report No. 76. Paris: Unesco/International Institute for Educational Planning, 1989.

Promoting utilization of evaluation results

There is considerable concern among evaluators (as well as among those who commission evaluation studies) about the non-utilization of evaluation results. It is said that too many evaluation reports are received, filed and forgotten, and that no use is made of their findings or recommendations.

There is some truth to the statement just made, and a part of the blame goes to evaluators. It happens too often for comfort that evaluation results are available long after the fact when the program or project is already an old story. But this may also be a case of a mild misunderstanding. Evaluators seem to think that there is knowledge-utilization if and only if decisions made by practitioners are clearly "knowledge-driven". There is, however, a less spectacular and more realistic view of knowledge-utilization. Evaluation results may often be used *without* formal acceptance of reports, and without formal credits and acknowledgements having been made to the evaluation studies and their authors. Indeed, evaluation studies may often change the structure of argument even of those who may be actively rejecting the evaluation results. This "utilization by rejection" is utilization nonetheless.

This view of indirect utilization should lead us to the obvious conclusion that to improve utilization, we must improve interaction with potential consumers from the very beginning of the evaluation effort. Evaluators should run an open ship, whereby participants can receive feedback as it emerges. Evaluators must also consider issuing interim reports which could be used to improve the program as and when new data become available.

A most important caution

Before going on to the next and last section of the handbook for a discussion of the politics of evaluation and the training of evaluators in the Third World, we must hark back to the earlier chapter on the management of evaluation. We had talked there of the methodological triangle of evaluation to include: MIS, NE and RE. While we have discussed each of these three methodological approaches in three separate Parts of the handbook, this is not to suggest, of course, that the three are separable. Indeed, in real-life settings, all the three approaches will be used within the same program, and sometimes within the same evaluation study. The need and the value of integration among these three approaches should never be lost sight of.

Things to do or think about

1. Examine the report of an evaluation study recently completed at your center, department, or ministry. Do you find it to be a complete and comprehensive report? How would you reorganize the report to make an improvement on the present version?

2. Was the above evaluation study timely for the practitioners of the program that it evaluated? What can you learn about the history of its utilization?

3. Prepare an oral presentation for a group of farmers based on an evaluation study done in your country on the subject of agricultural innovation.

Part VI

Some Important Related Concerns

This part of the book discusses the politics of evaluation and the need to establish evaluation standards for meta-evaluations. Another important related question, that of the training of evaluators, is also presented. It is divided into three chapters:

17. Politics of Evaluation, Ethics and Standards
18. Conducting Evaluation Training in the Third World, and
19. Conclusions.

CHAPTER 17

POLITICS OF EVALUATION, ETHICS AND STANDARDS

Information is power. Information can be put to political uses. Hence, evaluation, which creates somewhat "objective" information on the effectiveness of literacy and development actions, has political implications. In order to establish this objectivity, technical and ethical standards need to be observed.

Handling the politics of evaluation

How can we handle the politics of evaluation? No sure-fire formulas can be taught. In any case, most of us who have worked (and survived) within bureaucracies are not all that naive about the politics of survival and advancement within bureaucracies. Each one of us is perhaps somewhat qualified already in the art of "file-manship" and even "one-up-manship"! Yet, some general suggestions for handling the politics of evaluation may be in order.

There are two aspects to the politics of evaluation: (a) the evaluator should not be punished for doing the evaluation which may be seen as having produced "embarrassing" information; and (b) the information produced by the evaluation study should be put to practical use. Political problems do arise when, on the one hand, the evaluator seeks to make too much capital out of the evaluation study; and, on the other hand, creates information that threatens the various stakeholders within the system. Without compromising one's personal and professional integrity, one can do things, however, which will cool the politics surrounding the evaluation study.

Defend your right to undertake evaluation

Defend your right to conduct the evaluation. Let people know that evaluating is an integral part of good literacy work. Quote from a presidential speech, from planning documents, or from published prospectuses or reports of the parent institution. Your institution is bound to have declared evaluation to be a necessary part of its mission, though no one may have paid much attention to this particular objective. In an educational setting (as distinguished from an administrative setting), the right to evaluate can be defended as part of your professional interest. You, as a professional, are supposed to have an interest in evaluation.

Keep a low profile

There is a need for an evaluator to keep a low profile and have a sense of modesty about the evaluation study done. The evaluator should not demand to be considered a star on the institutional horizon. The report should be presented without too much fanfare, as a matter-of-fact collection of feedback information on the program. It should not be touted as a breakthrough of some sort.

Provide a framework of expectations for evaluation results

No program will ever be found to be performing at 100% efficiency level. Especially in social change programs, participation levels of as low as 30% may sometimes be acceptable. Before presenting the feedback on performance of a program, one must indicate what would be a reasonable level of expectation of performance. Findings should then be presented within such a framework. In other words, the readers of an evaluation report should be provided with standards and yardsticks with which to judge the success or failure of a literacy program or a development action. Without norms, readers may not know whether to be satisfied or to be dissatisfied with a particular set of results.

As we have said elsewhere, the focus should be on finding causes, not culprits. This is not to say that the program staff is never at fault and that as evaluators we should be finding alibis for them. Yet, processes and personnel must not be confused in the allocations of credit and blame. Things must be kept in balance.

Begin with a "draft" report

An important part of the political strategy may be to present the evaluation report to colleagues in a "draft" form, offering to do a final draft on the basis of collegial discussion and review. In a revision that follows, it will be important to neutralize the politics but *without* compromising the integrity of results.

Indicate possible actions

Indicate the actions that must be taken to make use of the findings of the study. Distinguish between things within the institution's control and those outside its control. Start with what the institution can do within its existing mandate -- such as curriculum revision, preparation of new testing procedures, etc. If the implementation of findings demands additional work, offer to do it singly, or with the help of a group or a committee. What we have suggested here may not always work, but it will increase the chances of an evaluation study influencing actions within the setting of a training center or a training institute.

Ethics of evaluation

The professionals are supposed to be self-disciplined, and professional institutions are meant to be self-regulating, normative subcultures. For that reason, ethical behavior has always been central to the lives of professional workers -- doctors, lawyers, accountants, teachers, engineers and, of course, researchers and evaluators.

In the U.S.A., the question of the ethics of the professions has come center-stage as politicians, bureaucrats, bankers, and ministers of God have all made a spectacle of their venality on national television. Of course, the U.S.A. cannot claim uniqueness for its lack of ethical standards in daily life. Ethical problems have indeed appeared worldwide.

In evaluation, questions of ethics emerge in different contexts. Ethical problems will be involved if:

1. the evaluation study is being undertaken to embarrass another individual, to kill a program, or to provide legitimacy for an action the politicians have already decided upon;
2. the evaluation data are being cooked up or if anti-social or criminal behavior is being encouraged or abetted so that the evaluator can collect the required data;
3. the privacy of the respondents is not protected, and respondents are being personally violated;
4. the data are being falsified during analysis to suit personal or political purposes; and

5. the results of an evaluation are withheld for selfish purposes.

It is not possible, of course, to ensure ethical behavior from evaluators. However, it is possible to discuss all the ethical dilemmas an evaluator is likely to face and to teach evaluators to learn to engage in ethical decision-making.

Evaluation standards: Evaluation of evaluations

Evaluators should themselves be held accountable. Their work must be judged according to some agreed standards of technical competence and ethics.

The Joint Committee on Standards for Educational Evaluations of the U.S.A. has developed 30 standards which the committee suggests should become the working philosophy of evaluators and should guide and govern the evaluation efforts of educators (and development workers).[1] A summary of these standards is provided below:

Summary of the standards for evaluations

A. *Utility standards*

Evaluation should serve practical information needs.

1 A(1) Audience identification
Audiences involved in or affected by evaluation should be identified.
2 A(2) Evaluator credibility
The evaluator should be both trustworthy and competent.
3 A(3) Information scope and selection
The scope and selection of information collected should enable pertinent questions to be answered.
4 A(4) Valuational interpretation
Value judgements used by evaluators should be made clear to readers.
5 A(5) Report clarity
Objectives, procedures used, findings, and recommendations should be clearly stated.
6 A(6) Report dissemination
Findings must be disseminated for use.

7 A(7) Report timeliness
Evaluation must be completed on time for use by decision-makers.

8 A(8) Evaluation impact
Evaluators should encourage follow-through by the concerned audiences.

B. Feasibility standards
Evaluation should be realistic, prudent, diplomatic and frugal.

9 B(1) Practical procedures
Procedures should be practical and should avoid disruptions of normal work.

10 B(2) Political viability
Evaluators should attract cooperation of various interest groups, avoid their attacks, ensure against misuse of results.

11 B(3) Cost effectiveness
Results should justify resources expended.

C. Propriety standards
Evaluation should be conducted legally and ethically and should contribute to human welfare.

12 C(1) Formal obligation
Formal obligations and contracts may be developed between various parties involved (especially in the case of external evaluations).

13 C(2) Conflict of interest
Should be avoided and where unavoidable should be dealt with openly and honestly.

14 C(3) Full and frank disclosure
Pertinent findings should be fully disclosed; limitations should be frankly stated.

15 C(4) Public's right to know
The public's right to know of evaluation results should be respected (unless it is clearly a matter of individual privacy or public safety).

16 C(5) Rights of human subjects
Rights of human subjects should be respected and protected.

17 C(6) Human interactions
In their interactions with subjects, evaluators should respect the dignity and worth of individuals.

18 C(7) Balanced reporting
The reporting should balance both strengths and weaknesses of what is evaluated.

19 C(8) Fiscal responsibility
Financial and other resources spent should be accounted for.

D. Accuracy standards

Evaluation should convey technically adequate information.

20 D(1) Object identification
What is being evaluated should be clearly identified.

21 D(2) Context analysis
Context of evaluation should be sufficiently described so that its influences on the object evaluated can be identified.

22 D(3) Description of purposes and procedures
The purposes and procedures of evaluation should be described in enough detail.

23 D(4) Defensible information sources
The sources of information should be described so that the reader can see if they are defensible sources.

24 D(5) Valid measurement
Evaluation instruments should be constructed and applied in ways to ensure validity.

25 D(6) Reliable measurement
Evaluation instruments should be constructed and applied in ways to ensure reliability.

26 D(7) Systematic data control
Data should be reviewed and corrected at various stages of the study.

27 D(8) Analysis of quantitative information
The analysis should be appropriate and systematic.

28 D(9) Analysis of qualitative information
The analysis should be appropriate and systematic.

29 D(10) Justified conclusions
Conclusions should be explicitly justified.
30 D(11) Objective reporting
The reporting should be objective and unbiased.

Some of these standards may seem too tough, and some too squeamish and overly fastidious, to evaluators working in cultures other than the United States where these standards were developed. Evaluators everywhere should, however, take these standards into account to the extent feasible.

Things to do or think about

1. What do you think of the practicality of suggestions made in the first part of this chapter for managing the politics of evaluation?

2. Evaluate a recent evaluation study done in your country in terms of the 30 standards for evaluation listed above.

Note

1. The Joint Committee on Standards for Educational Evaluation. *Standards for Evaluations of Educational Programs, Projects and Materials.* New York, NY: McGraw-Hill, 1981.

CHAPTER 18

CONDUCTING EVALUATION TRAINING IN THE THIRD WORLD

The training of evaluators is a challenge in any context, but in the Third World environment, evaluation training presents special problems. Local training capacity is almost non-existent. Typically, evaluation training comes to these countries through outsiders, and quite often within the framework of technical assistance. The Action Training Model (ATM) discussed here was first developed for the delivery of evaluation training to literacy and development workers in Kenya during 1979-82. Since then it has been tested in a variety of training settings in many different countries, and it is now presented as a model of choice for conducting training programs for evaluation personnel. The ATM has often been adapted for in-country use by Third World trainers of evaluation and for development training in general. At other times, selected components of the model have been incorporated by trainers in their training programs.

This book was conceptualized and written as part of a particular training approach concretized in the "Action Training Model" (ATM).[1] Of course, the book is by no means exclusively tied to the model, but will surely be read to serve multiple objectives, and will be used in varied training settings.

The Action Training Model was first developed and tested in the context of a series of workshops on evaluation in Kenya during 1979-82. The ATM has since been used in other training settings to train curriculum developers, writers of materials for new literates, and in workshops to produce distance education materials in Botswana, Kenya, Malawi, and Zimbabwe during 1979-90 under the aegis of the German Foundation for International Development. Direct experience with the model as well as its systematic evaluations have pointed to the ATM's effectiveness.[2] We can therefore recommend it as an effective model of training middle-level personnel in development settings.

While the ATM was developed within the international context of technical assistance, it has been used with equal effectiveness in intra-national settings. In conditions where the model could not be used in full because of want of resources, or of lack of total acceptance by everyone involved, different components of the ATM have been used in the delivery of training with most satisfying results.

The training of evaluators is difficult under any circumstances.[3] It is particularly difficult in Third World settings. Institutions of higher education in the Third World seldom have the resources to offer professional evaluation training either to their students in residence or to practitioners already at work in the economy. Government departments of education or special institutes for development training are similarly unprepared to offer such training to their program staff. Matters are not helped at all by the fact that the initial pool of people with a general research background, who could adapt their methodological skills to evaluation is quite small. Finally, institutional norms and expectations are generally unsupportive of evaluation. The tolerance for evaluative information that may bring bad news for some people and programs is low in the political culture of most nations.

The Action Training Model

The ATM was designed to overcome the problems just listed. It did not, of course, arise complete and whole out of nowhere. We learned from our experience and we were in a continuous process of retooling and refining the model over months and years.

It should also be indicated here that the introduction of the ATM to new training cultures will not be without problems. While almost everyone -- planners, resource persons and trainees -- would agree with the approach at the level of rhetoric, there will be much stalling when it comes to implementation. Certainty is comfortable and uncertainty creates anxiety -- for everyone. It is much more comforting to have stated training objectives, well-designed timetables, accompanying lecture notes already written and duplicated, even when the objectives adopted may be irrelevant, the whole of the materials may be academic and no learning may take place at the training site. Dealing with a living training system with real objectives, concrete learning needs and particular information demands is not only full of uncertainties but is much more challenging for everyone involved. Planners are afraid of losing control and not being able to sanction and approve of what will happen at the training site. Resource persons at training sites are afraid of faltering and of being exposed. Learners are afraid of taking responsibility for their own learning.

Analytical descriptions of the ATM and evaluations of its implementation have appeared elsewhere in the literature.[4] Here we shall describe the model in more or less chronological steps and in "user-friendly" terms.

The ideology of technical assistance and the philosophy of knowledge transfer through training

Long-term commitments

Long-term commitments were made and expected. This meant long-term commitments from donors, from host country institution(s), from workshop faculty and resource persons and, of course, from participants. There was no legal contract in most cases, but there were expectations of long-term commitment, which were in most cases fulfilled.

International, regional, and national projects

All kinds of program and project -- international, regional, national and sub-regional -- have their place in technical assistance. Training of middle-level personnel works best, however, in national contexts. National needs can be defined more clearly. A large enough number of people can be trained to give a country a critical mass of trained manpower in a new professional sector. Costs can be kept low since national travel costs will be lower and international rates of travel and subsistence allowance will not apply in most cases.

Transfer of responsibility

A transfer of responsibility to the host country's professionals should, of course, be an important consideration of all technical assistance, including that involving knowledge transfer through training. There is no justification for continued dependency on a team of outsiders. Since responsibility can only be transferred to professionals in the host country capable of accepting it, it requires the professional capacitation of the host country's faculty and resource persons.

Such training and orientation should be an important part of the training project.

Institutionalization of initiatives
Skills learned through training and local capacitation are vitiated if there are no structures to "contain" the skills and capacities learned. These structures can be official or non-official, that is, voluntary. In either case there is the need to institutionalize -- to integrate the program within the on-going programs of an appropriate host country institution and the commitment of local resources. Both the ownership and the responsibility should shift to the host country in due course of time.

Generative interventions
Generative interventions are fertile. They have consequences for a multiplicity of interconnected systems over a period of time. Evaluation is an inherently generative process. It is indeed the obverse of "planning" and it is directly interfaced with management and implementation. The generative aspects of evaluation training should not, however, be left to emerge in the minds of trainees, but should be explicitly pointed out.

Collaborative planning
Collaboration is the key in all planning and training. It is ideologically necessary, for people must take control of their own training and socialization. But collaboration is also functionally wise, for the local people have information about their communities and cultures that is simply unavailable to the outside planner or trainer. Also, people are most motivated to do things they have had a hand in planning, designing and implementing. Collaborative strategies should be applicable in all settings of decision-making, learning and evaluating at all the different levels.

Mutual obligations
The mutual obligations of trainees and trainers need to be clearly stated. Trainers, whether from the outside or the inside, should not try to "buy" participants for their training programs and projects.

In the Kenya evaluation series, once participants were there, they were expected to work. If a participant was unable to complete the field work on the proposal between the first workshop and the review panel (or the next workshop), he or she was told not to return. In some workshops, a trainee or two were sent away from the workshop site because they had come without completed work.

Minimum fanfare

Workshops under the ATM were working workshops. There was a minimum of fanfare. Opening and closing ceremonies were held only when it suited the obvious need of providing national visibility to a particular issue or when certain leaders or institutions were to be brought aboard.

Internal evaluation

The material in this book can be used to conduct evaluations large and small, internal and external. The ATM, however, is built on the assumptions of internal evaluation. We believe that evaluation is an instrument of the literacy professional, not of the educational police, and that it is to find causes not culprits, reasons not renegades. While external evaluation will continue to be conducted by outsiders to serve their own policy interests and needs for resource allocations, we believe that middle-level development people should be trained to conduct evaluations of their own program, on their own, for continuous feedback on how their programs are succeeding and where they might be failing.

Modelling time and effort

The ATM expands time. Everything does not have to be done within the confines of a two-week workshop. People can go back home and work on their own. Indeed, the training approach developed involved a training cycle of approximately one year's duration, composed of two two-week workshops (A1 and A2), with a panel (Pa) in the middle. A second cycle of two workshops (B1 and B2) and a panel (Pb) would overlap with the cycle A as follows:

(A1)...(3-4 months)...(Pa)...(3-4 months)...[(A2)(B1)]...
(3-4 months)...(Pb)...[(B2)(C1)], etc.

Each of the two periods of three to four months' duration was used systematically as part of the training cycle. (Panels later on became full-fledged workshops.) These are the periods for learning by doing -- they provide the time for action. Indeed, it is from this feature that the Action Training Model gets its name. The model is so called because it demands *action* from trainees in the application of skills learned during training, in their own work in real-life institutional settings.

Overall project planning

Project description

Writing a project description was the first step in planning a workshop (or workshop series as appropriate). The project description became the instrument of discussion and communication among everyone involved. It was used to explain the project to bureaucrats in the ministries, to host institutions, to sponsors, potential participants and whoever else was interested. The project description contained much of the information now included in this chapter on training.

Choosing an institutional home

We have already talked about the need for institutionalization of outside initiatives. The choice of an institutional home for evaluation training will depend on the context of the country. A university setting may be ideal, but a training institution delivering development training in literacy, health education or family planning could serve as well. A government department could serve equally well. Nongovernmental organizations such as the Indian Adult Education Association in New Delhi should also be all right.

Orientation of local faculty

Proper orientation of local faculty and resource persons is essential. This should involve more than a cursory introduction by way of this book. The ideology and philosophy of the ATM should be fully explained. The present chapter should be read carefully and supplemented with materials included in the "Notes" to this chapter.

Working with the ATM indeed requires a new way of doing things, if not a new socialization for workshopping. The ATM assumes a learning community involving all the trainees and all the trainers. Trainers do not simply come, deliver their lectures at their appointed hour, and leave. They are expected to be part of the "living system", which means that they are expected to be present all the time, taking responsibility for everything that happens. It is not always clear who will be asked to do what and at what time. In one particular training session, different facilitators may contribute. There is no such thing as an interruption.

An evaluation resource center

Professional libraries in the Third World are abysmally poor. Few libraries have books on evaluation in their collections. It is important that as part of the institutionalization of the initiative, a small evaluation resource center be established within the library of the selected institution. This collection should travel to various workshop sites as needed.

National Evaluation Group

Those serving as faculty at the workshop(s) and some others should be selected to form a National Evaluation Group (NEG) to help workshop participants with their projects by mail, by telephone and by personal visits. Such a NEG should, at the appropriate time, form the nucleus of a National or Regional Evaluation Association.

Contact with sponsoring agencies

As part of pre-planning, contacts should be established by trainers with institutions that will sponsor trainees for the workshops. Contacts should involve more than literacy institutions. Indeed, all development institutions should be covered. Preference may be given to "training institutions" in order to multiply the effects. Sponsoring institutions should be given a clear idea of the benefits, but also of the depth of commitment required on their part.

Preparation for a specific workshop or workshop series

Faculty recruitment and orientation

A core faculty may come from the host institution, but the larger faculty group should have a national representation. Separate resources should be assigned to faculty orientation and long-term professional development. As indicated earlier, the special ways of doing things under the ATM should be made clear.

It should be remembered that while faculty would agree to the ATM methodology at the intellectual level, they would resist it in practice. Delivering prepared lectures to a group with few questions asked is a much more comfortable position for the lecturer than the demands put on an instructor in the context of the ATM.

Participants' recruitment

Institutions should not simply be asked to send delegates to a workshop. There should be a combination of individual merit and interest on the one hand, and institutional commitment on the other. This means that individuals should apply, the workshop should select, and institutions should agree to release participants as if they were on official duty. Selections should be preceded in all cases by interviews at the applicant's site of work. The kind of commitment required of participants should be made clear to everyone concerned.

Choice of site

The site should be in the nature of a retreat with a minimum of discomfort and interference. If possible, participants should be discouraged from bringing their own cars to the workshop site.

Needs assessment

A generalized needs assessment should be carried out, *not* to design a curriculum but to develop a concept of what the needs might be and to conceptualize a "tentative curriculum." One may find oneself jotting down topics such as:

- What is evaluation?
- Evaluation planning and management.
- Writing evaluation proposals.
- Item writing.
- Evaluation of primers.
- Report writing, etc.

Some groups may want training in MIS design and use. Others may want to be able to handle NE. In any case, we are talking here of soft focusing. We do not want to pre-empt the honest effort to re-invent the workshop curriculum in the local setting, in collaboration with the real group of participants.

On site, during the workshop

Preparation of the site

The site should be comfortable and congenial. Training facilities should include a large hall with a lot of wall space. It should be possible to seat 35-40 people in this hall in a horse-shoe arrangement. *In other words, participants should be seated as at a round table rather than in formal school rows.* At least four rooms should be available for group work, and an additional room to serve as the secretariat. There should be an adjacent lounge where faculty can have one-to-one consultations. All areas, including bedrooms for participants, should be well-lighted. Finally, there should be a social room with radio and newspapers.

Reception and registration

It is important that reception and registration of participants be properly handled. They must feel welcome and integrated.

Resources on site

There should be enough instructional resources on site: books for distribution among the participants and resource persons; writing pads; pencils; foot rulers; scissors, etc. The mobile library of books on evaluation should be brought to the site.

The first day of the workshop

The first day of the workshop is the most important day of the workshop. It should begin with a plenary session. While it is important that plenary sessions be well led, people can sometimes go overboard by having a chairman, a moderator and a speaker! Such arrangements can become somewhat absurd and a lot of time can be wasted. As far as possible, the speaker should conduct the session unless he or she asks for someone else's help.

Building a team among faculty and a learning community that encompasses every participant is absolutely essential. Self-introductions should be the first order of business. Participants should fill a form, giving personal data. This they should first use to make their personal introductions, and they should later give it to the workshop director for the preparation of a list of participants. The process should not be unnecessarily hastened. Be patient. It is always time well spent.

In a second round ask each and every one of the participants to indicate what they have come to learn at the workshop. Patiently write key words representing all interests on sheets of newsprint taped on the walls. Once everyone has had an opportunity to express their interests, cluster and sequence learning needs.

Tell them how the participants' interests will be tackled *as far as possible* in the workshop program. Tell them why you do not have a ready-made timetable, but are ready to present the particular version of the workshop they need.

Tell them more about the ATM, pointing out that they will learn more about it as the workshop unfolds.

Explain the role of a steering committee and establish such a committee. It should include 5-7 people, at least two of whom should be participants. Participants may serve by turns. Explain what will be done in steering committee meetings, and why.

Distribute workshop materials. Then "walk through" each and every item. In other words, make participants familiar with what the various materials include and how they might be used during and after the workshop.

Talk of the various instructional approaches that will be used in the workshop -- the plenary, the group, the consultation between individuals, and individual work within group settings.

Finally, help the group learn "the language of discourse", that is, the meaning and definitions of evaluation terms they must know in order to start working. This should be more than enough for the first day of the workshop. If some of it, it seems, will spill into the first part of the next day, do not panic.

The first steering committee meeting

In the first steering committee meeting, the workshop director as temporary chair should recount for the group the role of the steering committee in the overall ATM. It should be pointed out that the steering committee is the instrument of formative evaluation of the workshop, a mechanism of collective reflection, and a mechanism of control of the workshop by the group rather than by the team of outside experts. The questions for the steering committee every evening will be the same: How was the day today? Are we all -- facilitators, staff, participants -- doing our very best? Are we all learning? Are participants fully involved? Is this what we had hoped would be achieved today? Where did we succeed and where did we fail? What could we have done differently? Knowing what we know now, what should be the tentative plans for tomorrow? Who should do what, when, taking how much time?

A steering committee meeting should not take longer than one hour. Unfortunately, some steering committee meetings have dragged on for two to three hours. Thus, a balance needs to be struck between democracy and effectiveness in the conduct of the steering committee. It may be opportune to have only experienced individuals to chair the meetings. Indeed, asking the local workshop coordinator to be the permanent chair may be a good idea to provide continuity to this important mechanism of planning and formative evaluation.

To avoid the task of having to schedule a steering committee every day, establish one time and one place for the steering committee for the total duration of the workshop.

The second day and later

Begin every day with a plenary, even if it is a short plenary session. Start with any administrative announcements. Then, present to the group the timetable of the day past, as it actually emerged. Give the group a summary of the deliberations of the steering committee of the previous evening. Remind the two participants who are on the list to attend the steering committee meeting that evening. Then continue with the program as developed in the last night's steering committee meeting and now approved by the participants.

The core instructional strategy of evaluation workshops

The core instructional strategy in the evaluation workshops should be to enable each and every participant to write an acceptable evaluation proposal to take back home to implement. Whatever else is done, the workshop should assist in (i) the development of an evaluation proposal, and (ii) the learning of concepts and skills that will enable the participants to develop future evaluation plans more or less on their own.

The essential instructional materials

There is a tendency on the part of the participants of workshops to collect papers and be fed on lecture notes and outlines. Lecturers are also part of the same workshop pathology. While it may be necessary to prepare and distribute some lecture notes and outlines, the tendency to re-do what is already in the workshop texts should be avoided. The special workshop manual should be put to work, and the time and resources of the workshop thus saved should be redirected to teaching and learning.

Expanded settings of instruction

At least four different types of instructional format are visualized: the plenary sessions where inputs are made or work of individuals and groups is offered to the total workshop; group work; individual consultations and individual work. *Individual work should be done in a group context.* That is, the individual may work individually, but may do so seated in a group setting so that the same resource person can reach all the different individuals sitting in a particular group setting.

Resource person and participant relationships

The ATM is not based on a set of lectures followed by polite discussion. It involves educational encounters between and among facilitators and participants involving provision of feedback to each other. In settings where participants belong to different races, nations, ethnic groups, and have varying social and bureaucratic status, provision of feedback has to be both useful and tactful. This is not so easy. However, when people have established initial relationships of trust they can say a lot to each other. The excitement of learning is worth the occasional tense moment.

Social architecture

The above is one of the reasons why due attention should be paid to the social architecture of the workshop. Some of the anticipated problems should be brought out in the open. Effective use should be made of official receptions and socials, which are always a part of national and international gatherings. A social on the third evening of a two-week workshop may be much more useful than a big party on the last day as a send-off.

Directing with a low profile: Day 3, 4, 5 and so on

The ATM is learner-centered, participatory, and flexible. But this does not mean that in ATM, the expert is marginal, outside the process of participation, and unwilling to lead for fear of imposing his or her opinions. Nothing could be farther from the truth. In being learner-centered, ATM allows the learner to participate in choices about what is learned and how, but that does not mean that the expert does not guide and teach. The ATM is participatory, but that does not mean that the expert does not participate and make the expert point of view known. The ATM is flexible, but it is the moral duty of the expert that flexibility does not go beyond reasonable limits and the training event does not flop.

Once the curriculum needs of the group have been expressed and made visual on posters on the wall, and once the steering committee is in business, it is the duty of the workshop director to hold the workshop together as a system. This will mean dialog with various stakeholders, discussion, reflection and more reflection.

The director should become the conscience-keeper of the workshop letting everyone know what he or she sees happening and why. The director should also become the workshop's time-keeper, making the group aware of time utilization and what changes might be necessary to be able to finish the evaluation proposal in time.

On the basis of earlier experience with similar workshops, and of contact with participants in the on-going workshop, the director should look at the expressed needs of participants and determine how much it is possible to do within the time left and how much may have to be handled in a second workshop or through other means at a distance.

The director should also work with the team of facilitators, knitting them into a team, helping them grow and even prepare for their inputs.

The day before the last day

Part of the day before the last day should be used to review what has been done and what remains, and to take any emergency measures that might be necessary to have a satisfactory resolution of the workshop experience.

The last day and a look ahead

The last day must include at least two items: an evaluation of the workshop experience; and a systematic look ahead. The evaluation of the evaluation training should be designed by a group including representatives from among the participants. Participants could be asked to contribute items for inclusion in the final evaluation. If at all possible, evaluation should be processed and feedback provided to the participants before departure from the workshop site.

Due time should be given to the look ahead. It should be made clear what should be done, when, and how. Participants should understand what should be done between now and the next workshop; where they can go for help; who will pay for what services rendered or obtained; etc. This further schedule should be developed as part of a collaborative effort and put in writing.

Notes

1. A description of the test-in-use of the ATM, written at the conclusion of the first phase of the Kenya project, and before the

transfer of responsibility to Kenyan colleagues in June 1982, is included in Bhola, H.S., *Action training model (ATM): An innovative approach to training literacy workers.* Paris: Unesco Unit for Cooperation with UNICEF, March 1983. A more systematic analysis of the assumptions and experiences with the ATM has been published recently: Bhola, H.S., "Training evaluators in the Third World: Implementation of the Action Training Model (ATM) in Kenya", *Evaluation and Program Planning*, Vol. 12, pp. 249-258, 1989.

2. Examples are:
   Müller, Josef, "Evaluation of basic education and development training programmes." Bonn: German Foundation for International Development, August 1980.
   Nturibi, Daudi N., "Experiences in training evaluators from training and development programs in Kenya, 1979-82." Bonn: German Foundation for International Development, January 1983.
   Mulusa, Tom, *Evaluation of basic education and development training programs: Mid-term evaluation of a workshop series.* Nairobi: College of Adult and Distance Education, University of Nairobi, April 1985.
   Gachuhi, D., Kenyi, C., and Matiru B., *Designing and Writing Distance Education Materials for Basic Education and Development Training Programmes* (Mid-term Evaluation 1985-1989). Bonn: German Foundation for International Development.

3. See Sechrest, Lee, (ed.), *Training program evaluators.* San Francisco: Jossey-Bass, 1980. Also, Davis, Barbara Gross, (ed.) *Teaching of evaluation across the disciplines.* San Francisco, CA: Jossey-Bass, 1986.

4. See Bhola, "Training Evaluators in the Third World," etc. (Note 1 above).

CHAPTER 19

CONCLUSIONS

A book cannot but be a personal statement by its author. Yet, we have sought to write a book in behalf of the professional community of educators and evaluators of adult literacy and nonformal education. We have taken clear, and some unique, positions. However, these positions have emerged from our experience in the practice of evaluation in Africa, Asia and Latin America. What we have proposed has not come from ideological or methodological dogmatism, but from the realities of "cultures of evaluation" as they prevail today around the world and particularly within the Third World.

While accepting the need of external evaluation for making inter-regional and international comparisons, we have emphasized internal evaluation. Thereby, we have accepted the view that the educator and the evaluator belong to the same ideal learning community.[1] Ideally, we have suggested, the role of the evaluator should not be separated from the role of the educator.

Since we expect *all* agents of education including grassroot workers to act as evaluators, we have had an important stake in demystifying evaluation. Some readers of the book may think that we may have made things seem simpler than they really are. To them we will only say that anyone who can be entrusted with the educational process could also be entrusted with the evaluation process. Having said that, we will make the further point that we need to engage in a continuous process of learning more as educators and evaluators, within the context of our particular programs of adult literacy and nonformal education. Demystification must be followed by the necessary evaluation training and learning of evaluation by doing.

This book has been about making "informed" decisions. There is the strong implication, throughout the book, that "evaluation" should be subsumed under "information generation". This conceptual step-up from the "concept of evaluation" to the "category of information" has both theoretical and practical reasons. The theoretical reason is, of course, that to evaluate is to produce information for decision-making. But program decisions need more than *evaluative* information. They need *descriptive* information as well. By keeping in mind the larger "category of information", we are able to understand the need of information available from sources other than occasional evaluation studies. The practical reason for the step-up from evaluation to information is even more important. When program improvement is discussed within the conceptual framework of "evaluation", practitioners too often draw an unintended conclusion: That what we need for effective program decisions are occasional evaluation studies. A confusion arises between the part and the whole. Instead of a whole "culture of evaluation" that accommodates both evaluative and descriptive information, they end up focusing on small parts of information generated by evaluation studies alone. Information generated by the program in the process of implementation, and contextual information available from outside sources, is lost sight of.

It should be recollected that we have suggested three approaches to information generation: (a) management information systems, (b) naturalistic evaluation, and (c) rationalistic evaluation. These together will enable campaigns, programs and projects of adult literacy and nonformal education to develop a dynamic "culture of evaluation". A paper-and-pencil MIS will be an indispensable source of descriptive information that can profile the size, scope and surface structure of a program of adult literacy. We see the MIS to be in a symbiotic relationship with *both* NE and RE. The numerical nature of an MIS should not mislead us into thinking that MIS and RE are congruent while MIS and NE are unconnected.

A beginning student of evaluation may have found parts of the book "heavy reading". Concepts such as "system", "paradigm", "reductionism", and some others may have remained unclear. The process of elaboration from concepts to indicators, to test items may have seemed simple in theory but difficult to practice. That should not be the reason to despair. That should indeed be taken as an invitation to read again and practice more. We do not expect all readers to accept everything that has been said in this book. If the book is used as a springboard to additional sources (particularly in languages other than English) and to alternative positions in the concept and method of evaluation, the purpose of this book will have been well served.

This book has not only discussed how evaluation should be holistically conceptualized,[2] but also how the *training* of evaluators should be conducted in non-academic settings in the Third World.

Here, again, our bias has been congenial to naturalistic approaches. We have talked of contextual training design, and of participatory learning of evaluation concepts and methods.

We end this book by remarking that evaluation is a matter of being at the same time intellectually disciplined, keenly perceptive, and wonderfully intuitive. It must also be said that evaluators have to be more than methodologically astute. Evaluators have to be moral.

Notes

1. Marshall, James and Peters, Michael. Evaluation and education: The ideal learning community. *Policy Sciences* 18 (1985), 263-288.

2. Chinapah, Vinayasum, and Miron, Gary. *Evaluating Educational Programmes and Projects: Holistic and Practical Considerations*. Paris: Unesco, 1990.

GLOSSARY OF TERMS

Action Training Model (ATM). A training model developed under the aegis of the German Foundation for International Development (DSE), Bonn, Federal Republic of Germany. The model emerged within the context of a series of workshops on the evaluation of basic education and development training programs. The model is so called for its emphasis on *action.* Trainees are required to make commitments to a full cycle of training experiences: first, a workshop where trainees learn generally about evaluation and develop specific proposals for evaluation studies; second, a mid-term panel where the trainees come with evaluation data collected by them during some six months of the implementation of their evaluation studies, review their experiences and prepare for data analysis; and, finally, another workshop where old trainees come back to report on their findings and new ones launch upon a new training cycle under the ATM.

Analysis of variance (ANOVA). A method of determining whether the differences between groups are statistically significant.

Attrition. Loss of subjects from a chosen sample during the course of a study.

Audit of an evaluation. Examination and verification by another independent team of the quality of an evaluation plan, the adequacy with which it was implemented, the accuracy of its findings, and the validity of its conclusions.

Base-line survey. An initial survey that can serve as a base for comparing changes observed subsequently.

Bias. A consistent alignment to one particular point of view which may make objective evaluation results improbable.

Case study. A detailed description and analysis of a single program, project, course or instructional material conducted within its educational or social context.

Code. To convert a given set of data or items into a set of quantitative or qualitative symbols. (Examples: 1, 2, 3 and 4; or L, M and H.)

Coefficient. A statistic (or value) that represents the degree of occurrence of a property or relationship. (Example: correlation coefficient.)

Concept analysis. The process of "unpacking" concepts to define them with such precision that they will have the maximally invariant meanings for most readers.

Content analysis. Identifying, categorizing and listing according to some rules, ideas, references, feelings or judgements found in a set of transcripts, documents, etc.

Context evaluation. Assessing and evaluating the environmental variables of a program.

Control group. A group which resembles an experimental group (the group which is subjected to a particular program or method) as closely as possible, but is not exposed to the program or method whose effect is being studied. It thus serves the comparative purposes of the evaluator.

Correlation. A statistic which indicates the degree of relationship (going together or happening together) between or among variables. Correlations can vary from -1.00 to +1.00.

Cost-benefit analysis. An assessment of the inputs and outputs of a program in terms of their monetary values.

Cost-effectiveness analysis. An assessment of the inputs, processes and outputs of a program in terms of the effectiveness of means employed for the ends obtained.

Criterion. A standard by which something is judged.

Criterion-referenced tests. Tests whose scores are interpreted according to the criteria of performance specifically defined by the teacher in regard to a particular group, and not by reference to performance of some comparable populations.

Data. Material gathered during the course of an evaluation study (both quantitative and qualitative) which is then used to develop information for decision-making.

Data analysis. The process of identifying ideas, themes, and hypotheses from the data, and the use of data to demonstrate support for them.

Data pieces. Individual tests, interview schedules, questionnaires and diaries that have been completed as part of the data collection phase of an evaluation study.

Dependent variable. A measure (for example, better nutritional habits) which is supposed to vary as a result of the introduction of an independent variable (for example, teaching of nutritional habits by the family health educator).

Design. A model or a clearly established set of procedures to determine how an evaluation study will be conducted. (See also *Training design.*)

Development. The processes that lead to greater production of wealth in a society and a just and equitable distribution of such wealth, accompanied by progressive consumption of education and culture, and commitments to universal brotherhood, peace and preservation of the globe.

Development training. Training of workers and change agents who will, in turn, impart economic, social and political skills to farmers, workers, housewives and youth to enable them to generate and sustain development within their societies.

Dissemination. The process of spreading information about evaluation objectives and results among those concerned with the evaluation study. The methods of dissemination may be written or oral.

Evaluation. Objective and systematic collection of information about a program, project, or instructional material for its improvement. (More recently in literature, evaluation is being defined as the "systematic investigation of the worth or merit of an object; e.g., a program, project, or instructional material".)

Evaluation system. An arrangement of methods, procedures and plans of action designed to provide decision-makers with information on the inputs, outputs, context and process of a given program.

External evaluation. Evaluation conducted by evaluators not on the staff of a program or project.

Extrapolate. To infer from what is known, something that is unknown. (Population figures for a country for the year 2000 may be extrapolated from the population growth figures during 1950-80.)

Feedback. A term borrowed from electronics: the return of part of the output of a system into the input for purposes of modification and control of the output. In the context of program planning, feedback means evaluative information on program effects.

Field test. A preliminary study of a program, project or instructional material in a setting very similar to the one in which it will be later implemented or used on a much larger scale.

Formative evaluation. Evaluation conducted during the very formation of a program, project or instructional material.

Generalizability. The extent to which claims and assertions made about a program, project or instructional material in one setting can be applied in other settings.

Goal-free evaluation. Evaluation of outcomes of programs and projects where the evaluator functions without knowledge of the purposes and goals of a program or project.

Human resource development (HRD). The education and training of manpower, both for formal and informal sectors of the economy, using both formal and nonformal systems of instruction.

Independent variable. A treatment variable introduced in an evaluation setting (example: a new teaching method), expected to create varying effects on a dependent variable (for example, performance on a test).

Indicator. Something that indicates, points, signifies; a gauge that represents another entity. Thus, a high drop-out rate in an adult education program may be an indicator of a lack of community motivation.

Input evaluation. Assessing the various resources used in conducting a program.

Institution building. The process of developing organizational arrangements or systems for the implementation of programs or projects on a long-term basis. (To institutionalize is to make a program more or less permanent through institution building.)

Instrument. An assessment device (test, questionnaire, interview schedule, or observation schedule) used for the purposes of evaluation.

Internal evaluation. An evaluation conducted by a staff member from within the organization whose program, project or instructional material is being evaluated.

Level of significance. A predetermined probability value which is used to decide whether the results of an evaluation study were really a consequence of of a program, project or instructional material, or whether they occured by chance. ($p = .01$ means that there is the probability of only one in a hundred for the program effect to have appeared by chance.)

Management information system (MIS). A system (computerized, manual or a mix of the computerized and the manual) including planning and implementation data in regard to a program or project. (See also *Monitoring*.)

Matching. The process by which subjects assigned to different groups are made to be as equivalent as possible. (Matching may be done on such variables as sex, age, education, socio-economic status, etc. A set of twins would be perfectly matched for the purposes of some studies.)

Mean. The sum of a group of scores divided by the number of scores.

Median. The score in a group of scores that is midway in the distribution.

Mode. The score in a group of scores that occurs most often.

Model. A design, description or analogy used to help visualize or make understandable something that is more complex.

Modus operandi analysis. A procedure similar to detective work whereby causes and effects are hypothesized, tested and analyzed to arrive at the most likely patterns of events and their consequences.

Monitoring. To monitor is to check on an on-going program or project for flaws or breakdowns, to enable decision-makers to regulate activities and to undertake corrective action. Monitoring is typically based in a management information system.

Naturalistic inquiry paradigm. Study of behavioral phenomena in natural settings and in their normal context, using methods drawn from ethnography, anthropology and sociological field studies. Also called the ethnographic or the phenomenological paradigm.

Needs assessment. The process of ascertaining the learning needs, health needs or other developmental needs of beneficiaries of educational and developmental programs. Needs assessments are a mix of "felt" needs expressed by beneficiaries and new needs "fashioned" by change agents.

No significant difference (nsd). A label which is used to say that the observed difference between two statistics could have occurred by chance. (See *Level of significance* above.)

Nonformal education. A collection of organized or semi-organized educational activities, operating outside the formal education system and meeting the immediate educational needs of both conventional and non-conventional learners. (Formal education is that which is provided by schools, colleges and universities. Informal education is that where neither the educator nor the one being educated is conscious of the process of teaching-learning taking place.)

Norm. A value or pattern of values representing the typical performance of a group or population.

Norm-referenced tests. See *Standardized tests.*

Objective-referenced tests. Tests whose scores are interpreted according to the objectives which a program, project or course was designed to teach, without comparing performance of other groups on the test.

Operational seminar. A training method developed within Unesco wherein participants experience on a reduced time-scale the total process of community work, problem diagnosis, needs assessment, field organization, materials design and evaluation in an actual field setting.

Output evaluation. Assessing the quality and quantity of the final product(s) of the program, also taking into account any unintended by-products of the program.

Paradigm. An example or pattern; a very clear example of an archetype. In evaluation, a paradigm is equivalent to the "intellectual ideology" of an evaluator.

Parameter. Any one of a set of properties whose value determines the characteristics or behavior of something.

Participative approaches. Designs, procedures and methods of planning, implementation and evaluation that are built upon the active involvement of the would-be beneficiaries of programs and projects.

Population. All the persons in the group to which the results from a study will apply. (Examples: all cotton farmers in the Lake Regions of Tanzania, all women of child-bearing age in Indiana.)

Post-test. A test to determine the effects of a program, project or instructional material after application or completion.

Pre-test. A test to determine level of performance before the start or application of a program, project or instructional material.

Problem complex. A whole set of interrelated problems (of planning, or of management, or of evaluation), emerging around a decision point within a system.

Process evaluation. Assessing procedural strategies and comparing effectiveness of different approaches to instruction, extension, animation and organization.

Product evaluation. Assessing the effectiveness of curricular or instructional products.

Qualitative data. Facts, claims and assertions in narrative form, and not in numbers. (Qualitative data can, however, be converted into numerical form by coding and scoring.)

Quantitative data. Facts, claims and assertions presented in numerical forms.

Quick appraisals. Quick evaluations, less comprehensive and less exhaustive than regular evaluations, conducted under conditions of emergency to investigate the cause of a breakdown, to anticipate problems or to get early returns on the impact of a program.

Random sample. A representative portion chosen from among the population; each individual in the population has an equal chance of being selected each time a selection is made.

Reliability. The property of an instrument giving the same reading or score when used by different investigators on the same entity, or by the same investigator repeatedly on the same entity.

Replication. The repeat of an evaluation study with all essential aspects of the study remaining unchanged.

Responsive evaluation. Evaluation that responds to the information needs of the various stakeholders in a program by providing evaluation feedback on concerns and issues raised by them, rather than evaluating what the evaluator thinks is worth evaluating.

Sample. A part of a population chosen according to some method to represent the total population.

Rationalistic inquiry paradigm. The approach borrowed from the hard sciences involving experimental design, randomized samples, controlled groups and statistical analysis.

Situation-specific strategy (3-S) model of evaluation. A five-step model that relates evaluation to change; requires the articulation of means and ends within an educational or a developmental program; proposes the development of profiles of information needs; suggests that situation-specific and strategic agendas for evaluation be developed; and that the choice of evaluation methodologies and techniques be both technically appropriate and practically feasible within the setting of evaluation.

Standard deviation (s). A measure of variability calculated on the basis of differences of individual scores within a group from the group mean. s-squared is called variance.

Standardized tests. Tests whose scores are interpreted in comparison with norms established in terms of some larger groups or populations.

Statistic. A summary number that describes the characteristic or property of a sample.

Statistical analysis. An examination of complex relationships between variables using empirical data and rules of statistics.

Statistics. The science of methods for analyzing data obtained from empirical observations to make descriptions or inferences. Thus, there is descriptive statistics, and there is inferential statistics.

Summative evaluation. Assessment of the impact of the total product, program, etc., comparing observed effects with anticipated or desired effects.

System. A whole emerging from an interacting and interdependent set of parts, subject to a common plan and having a common purpose.

Systems model. A model that looks at social reality as a system that can always be described in terms of inputs, processes, outputs and context. (See also *Model* and *System.*)

Taxonomy. An orderly classification that has some theoretical underpinnings.

Thick description. Detailed and faithful descriptions in the form of photographic records and protocols or written case studies.

Training design. A model or a clearly established set of procedures to develop a training program, involving planned selection of educational objectives, learner characteristics, teaching methodologies and learning environments.

Triangulation. Comparing and testing results from two or more different approaches to the solution of the same problem.

Unit of analysis. The social unit such as individual, husband-wife dyad, family, group, organization or community which is the focus of interest for the evaluator; which will determine the organization of data; and about whose behavior statements, claims and assertions will be made.

Unobtrusive measures. Methods of examination in which the evaluators do not materially interfere in the situation, but rely on indirect procedures to gather data.

Validity. The property of an instrument which is able to measure what it was supposed to measure.

Variable. A characteristic that can take on different values.

Variance. A measure of variability calculated on the basis of differences of individual scores within a group from the group mean. The square root of variance gives the value of standard deviation (s).